The understanding of generalised arithmetic (algebra) by secondary school.

Kuchemann, Dietmar

The Understanding of Generalised Arithmetic (Algebra) by Secondary School Children

Dietmar Erich Küchemann

Submitted for the Degree of Ph D from Chelsea College,
University of London, 1980

Abstract

Küchemann DE: 1980, The Understanding of Generalised Arithmetic (Algebra) by Secondary School Children.

A class test, Algebra, was developed as part of the research effort of the mathematics wing of the programme "Concepts in Secondary school Mathematics and Science". The test was developed in the light of interviews with 27 children and written trials involving 11 classes. In the summer of 1976, the final version of the test was then given to representative samples of 1128 2nd year, 961 3rd year, and 731 4th year children (13, 14, and 15 year olds) in English secondary schools (together with an unrepresentative sample of 103 5th year children).

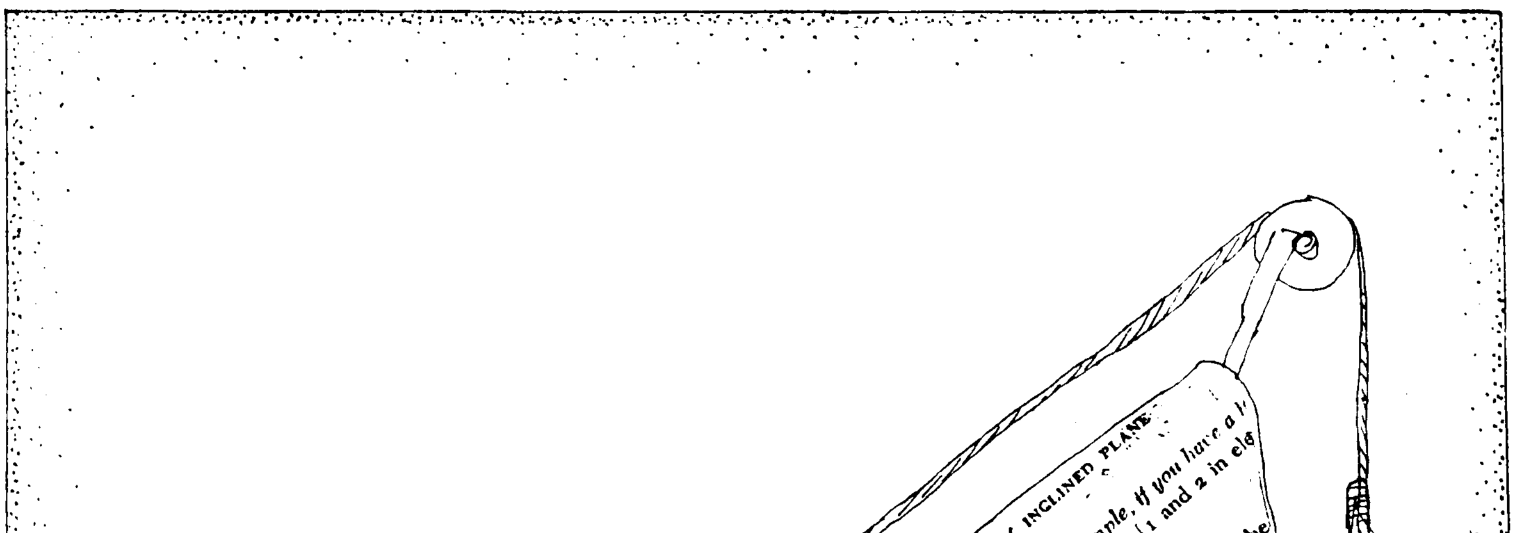
The test was designed to investigate the different ways in which children interpret the letters in generalised arithmetic: six ways were identified, which were called Letter Evaluated, Letter Not Used, Letter as Object, Letter as Specific Unknown, Letter as Generalised Number, and Letter as Variable. Subsequently selected items from the test were classified into four "levels of understanding". The items were selected by using statistical methods ("spider diagrams" and factor analysis) based on the correlation coefficient phi. Other coefficients were also investigated, as was a method based on characteristic curves and Guttman scalogram analysis.

Performance on the test was compared with performance on on other CSMS mathematics tests and also with a Piagetian class task devised by the science wing of CSMS. The latter comparison was used to classify the items on the test into Piagetian substages.

Acknowledgements

I would like to thank my friends on the CSMS programme and my supervisor, Professor David C Johnson, for their help and encouragement, and my immediate family, Maeve, Nancy and Alice, for their immense restraint, and also my mother for the loan of the Continental, which types ü's.

TO PIGLET



WITH MIXED FEELINGS

CONTENTS

Appendices

List of Tables and Other Figures

Chapter 1 INTRODUCTION

This thesis reports an investigation of children's understanding of generalised arithmetic that formed part of the research effort of the CSMS programme (Concepts in Secondary school Mathematics and Science) which was based at the Centre for Science Education, Chelsea College and funded by the SSRC from September 1974 - 1979. In particular, the the thesis discusses the development of a class test, Algebra, and the results obtained when the test was given to large, representative samples of 2nd, 3rd and 4th year children (13, 14 and 15 year olds) in English secondary schools. In what follows, the term algebra will be used in the restricted sense of generalised arithmetic, ie the use of letters for numbers.

Put briefly, the aim of the CSMS programme was to "help teachers and curriculum developers in the selection and presentation of appropriate materials and assessment of children's capabilities and progress" (Proposal to SSRC; CSMS,1973,p7). The approach adopted by the mathematics wing of CSMS was to develop a number of class tests, in the light of interviews with individual children, which could then be used both as instruments to assess children's levels of understanding in different areas of mathematics and to determine the cognitive demand of different mathematical tasks. Twelve tests were developed, each of which is described at length in CSMS (1981) and which are being made available to teachers and researchers through the NFER.

It had originally been intended to adopt a Piagetian framework for the research, both in the sense of applying Piagetian theory to the tasks outlined above and of extending or modifying the theory in the light of the CSMS findings. The science wing of CSMS kept very close to this aim, using in particular the descriptions of formal operations in Inhelder and Piaget (1958) as a framework for analysing science curricula and for describing children's conceptual level; they also examined the consistency of these descriptions by comparing children's performances on a number of class tasks developed in the light of the descriptions. On the other hand, the mathematics wing gradually moved away from a specifically Piagetian standpoint, partly because it was felt that the number of Piagetian studies of direct relevance to secondary school mathematics was insufficient; also, given the limited life of the research programme, it was generally felt that the task of devising an effective Piagetian scheme for analysing mathematics curricula would be too complex, given the numerous dimensions that in practice might affect the cognitive demand of mathematical tasks.

As far as the Algebra research in particular was concerned, a link was established empirically with Piagetian theory (see Chapter 12); also, both in the construction and analysis of the Algebra test, a Piagetian framework was adhered to in as much as it was assumed, at least as a first approximation, that children went through the same stages in their grasp of a given concept and that their level of understanding was consistent across different concepts.

The approach adopted for the Algebra and the other mathematics research was also strongly influenced by Piagetian theory in the sense that Piaget's system "puts the focus on pupils' thinking as shown by their explanations and justifications " and "gives recognition to the active construction of knowledge by each pupil" (Lovell,1980,p2). Thus rather than concentrating on the logical structure between mathematical concepts and tasks, the emphasis was placed on identifying the methods that children actually used. Initially, children were examined by means of individual interviews, and it quickly became apparent that the methods they used were often very different from what had been predicted, and from what they had been taught in school.

Class tests were developed in the light of the interviews and here marking schemes were devised that coded a range of responses rather than just classifying answers as right or wrong. By shifting from interviews to class tests, which could be given to large samples of children, it was hoped that the complex task of "identifying conceptual structures underlying school mathematics" (CSMS,1973) could be reduced to managable proportions. Also, it was hoped that the tests would provide teachers with easily used but sufficiently accurate instruments for assessing their own pupils' levels of understanding, as well as being used by the CSMS mathematics team to establish age-norms for children's levels of understanding which would provide teachers and curriculum developers with more general information about "the selection and presentation of appropriate materials" (ibid).

The thesis is divided into three parts: Part A discusses the development of the Algebra test and provides an interpretation of the results; the statistical methods* that were considered and finally used are discussed in Part B, and the relationship of the Algebra results to other research is discussed in Part C.

*Much of the statistical analysis was carried out by computer, and I would like to thank M. McCartney for looking after most of the programming and for his expert advice.

Chapter 2 THE DEVELOPMENT OF THE ALGEBRA TEST

The test had a gestation period of about 9 months. The first draft was written in June 1975 and the final version in March the following year. This version was then given to large representative samples of children in English secondary schools, with most children being tested at the start of the summer term 1976 (see Chapter 3). Development of the test began as a joint effort of the then members of the CSMS mathematics team (M Brown, BH Blakeley, DE Küchemann) but after a few months sole responsibility passed to Küchemann.

Development went through two major phases: on the first drafts of the test children were interviewed individually, and then versions were produced that could be given to whole classes of children at a time. In all, 27 children were interviewed (June - October 1975) and 13 classes tested (October 1975 - February 1976). Three drafts were produced during the interview phase and another 7 during the phase of class testing.

Initially it was far from certain how the test would develop. It had been decided to focus on generalised arithmetic rather than algebraic structure but a decision still had to be made about which aspects to investigate since it would be impossible to cover comprehensively even the most common aspects of secondary school algebra in one test. Little was known about how children coped with generalised arithmetic nor did a sufficiently articulated framework exist for interpreting their difficulties, although at that time there were some attempts to produce such a framework that looked promising.

Malpas and Brown (1974) had put forward the notion of "concrete" and "formal" models, which they derived from Piagetian theory, in particular from the account of adolescent thinking in Inhelder and Piaget (1958), and which they used quite successfully to determine the cognitive demand of O-level mathematics items and of certain teaching sequences in the SMP lettered-series texts (Malpas, 1974); more specifically some interesting studies by Collis (eg 1974, 1975a, 1975b) were just coming to light, in which the attempt was made to interpret children's performance on algebra items in Piagetian terms. The first draft of the Algebra test used some of these items and other items were constructed in the light of Collis's ideas; however, it should be said that most of the items were constructed on a far more intuitive basis, as items that seemed in some way typical of secondary school mathematics courses and that would hopefully elicit interesting responses, even if it was not quite clear what kind of responses to expect. Finally at this time the CSMS mathematics team were given access to results obtained by the NFER on some of their TAMS generalised arithmetic tests, which were being developed as a pilot study for what was later to become the APU (Assessment of Performance Unit). An analysis was made of these results (mostly by M Brown) to see whether factors having a substantial effect on item difficulty could be identified.

The Work of Collis

The aspect of Collis's work that was to have the most marked effect on the way the Algebra test was developed and later analysed was an investigation into equation solving undertaken by Collis in 1974 while on study leave at the University of Nottingham (Collis 1975b, pages 17 - 48). The investigation was concerned with the relative ease with which children solved simultaneous equations that required "substituting" (as in the case of the first item shown below) or only "matching".

Fig 2.1 Items requiring substituting and matching respectively.

Find the relationship between x and y

if	$3x = a$	if	$x + 5 = b$
and	$a + 3y = 180$	and	$y + 5 = b$

What was of interest here was not so much the distinction between substituting and matching, as Collis's observation that children worked with and interpreted the letters in generalised arithmetic in different ways. This notion was to become the focal point of the Algebra research, leading eventually to the formation of six categories for describing children's interpretations (see page 37). The interpretations identified by Collis are summarised in the table below, which is taken from a larger table made at the time the test was being developed.

Fig 2.2 Children's interpretations of the letters in generalised arithmetic (based on Collis,1975b).

Level	General Description	Responses to Specific Items
Level 1 about age 10	Children map letter directly into a specific number that seems viable at the time; stop if this does not work.	Their strategy is adequate for: $a + b = b + a$, True or False? $a + b + c = c + a + b$, True or False? Find $x \leftrightarrow y$ if $x + 5 = b$ and $y + 5 = b$
Level 2 about age 12+	Willing to map several numbers onto the letter in turn -a "guessing and testing" technique- but with the aim of finding THE correct number.	Several trials would appear to give more chances of success that the Level 1 strategy, but the more information that is obtained the more difficult it becomes to focus on the result that is required. eg, for Find $x \leftrightarrow y$ if $x + a + b = 180$ and $y + a + b = 180$, one child, after substituting several numbers for a and b says "It looks as though $x = y$" but then writes " $x + y = ?$" and stops, clearly puzzled.

Level 3 about 14-15	Seem to have extracted a concept of GENERALISED NUMBER, by which a symbol can be regarded as an entity in its own right but having the same properties as any number with which they have had previous experience.	Can solve Find x+y if y = b and r + 2b = 90 not by substituting specific numbers but by arguing "b is a number; 2b is twice that number and thus twice y; y and 2y makes 3y and thus y=30". More processing space is available for the final deduction because thinking is not cluttered with numerous specific examples.
Level 4	Letter interpreted as a VARIABLE.	Can cope with m + n + q = m + p + q is always? sometimes? never? true, which involves not only recognising that p and q both represent a range of values (are generalised numbers) but conceiving the remote possibility that the values of p and q may meet on any one element.

NOTE: the terms Generalised Number and Variable will be used somewhat differently later in this thesis.

Collis describes Level 2 as "late concrete-operational" and Level 4 as "formal-operational", which in Piagetian terms are equivalent to the early formal (3A) and late formal (3B) substages respectively (eg Collis, 1974).

Another interesting aspect of Collis's work is the notion of "acceptance of lack of closure" or ALC (Collis, 1972; Lunzer, 1973). ALC refers to the degree to which children are able to regard the outcome of an operation or series of operations as unique and meaningful (Collis,1974). To children at the lowest operational level (age about 7 year) expressions like 2+3 are meaningful in the sense that they can relate the elements to physical reality (eg 2 marbles and 3 marbles) and can imagine actually replacing the numbers 2 and 3 by the number 5. However it is not until the age of about 10 years that children can work with expressions involving numbers beyond their empirically verifyable age, eg 273+472, in the sense of being able to regard the outcome as unique without having to make the actual replacement to confirm this. Later still, at the stage of "late concrete operations", or what Collis also calls "concrete generalisations", they can treat expressions involving letters in the same way, where the letters themselves are regarded as unknown but unique numbers. Finally, at the age of about 15 years onwards children are no longer tied to the notion that numerical or algebraic expressions give unique results. This means that they are able to consider possible relationships between the elements in an algebraic expression, by examining what happens as the values of the elements vary. For example, a formula like V=LxBxH is no longer seen simply as device for determining V for given values of L,B and H, or even as a summary or generalisation of the values of V corresponding to discrete sets of values of L, B and H; instead, these children can meaningfully discuss questions like "What might happen to V if L is increased, B decreased and H held constant?".

The Interviews

Draft 1 of the Algebra test (see Appendix 2.1) contained 9 items based on or taken directly from Collis (1975b) (item 6v, which was designed to examine the notion discussed immediately above, of what happens when the values of unknowns vary, and Questions 8, 9 and 10). In addition, 7 longer questions were devised by the CSMS mathematics team, in which a given part of a question was in most cases dependent on the preceding parts (Questions 1 to 7). Drafts 2 and 3 were similar.

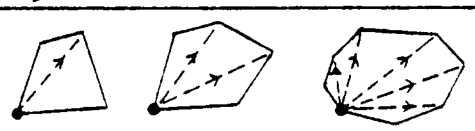
In the light of Collis's ideas and the individual interviews, an attempt was made to classify and order children's responses to each item on Drafts 1, 2 and 3. The tables below were devised in the summer of 1975 and show this classification for Questions 2, 3, 5, 6, 8 and 9 of Draft 1, after 18 children had been interviewed.

Fig 2.3 Classification of interview responses.

Question 2

RULE SEEKING	RULE USE	Verbal RULE ARTICULATION	Algebraic RULE ARTICULATION
$1 \longrightarrow 3$ $2 \longrightarrow 6$ $3 \longrightarrow 9$ $4 \longrightarrow$: $8 \longrightarrow$ $\longrightarrow 30$ $300 \longrightarrow$		What is the rule?	Can you write it as $n \longrightarrow$?
a. Uses limited information, eg if $3 \longrightarrow 9$, then $4 \rightarrow 16$.	a. Confused about rule.	a. Unable to express rule verbally	a. Unable to cope with n.
b. Uses number pattern but not looking for operation (ie working down rather than across)	b. Confused about inverse.	b. Confused about rule.	b. Regards n as "any number" so maps n onto one "any number". b'. Uses letter pattern, eg $n \longrightarrow p$.
c. OK.	c. OK.	c. OK.	c. Regards n as "any number" and maps n onto one multiple of 3. d. Maps n onto "any" multiple of 3 by giving one as just an example, or by giving a list. e. OK.

Question 3



RULE SEEKING	RULE USE	Verbal RULE ARTICULATION	Algebraic RULE ARTICULATION
$4 \longrightarrow 1$ $5 \longrightarrow 2$ $8 \longrightarrow 5$ $10 \longrightarrow$	$\longrightarrow 125$	You get the number of diagonals by	For a shape with k sides, you can draw .... diagonals.
a. Ignores given information, uses drawing.	a. Confused about rule.	a. Unable to express rule verbally.	a. Unable to cope with k, or regards as actual figure.
b. Uses limited information, eg if $5 \longrightarrow 2$, then $10 \longrightarrow 4$.	b. Confused about inverse.	b. Confused about rule.	b. Maps k onto one "any number". b'. Uses letter pattern: $k \longrightarrow h$. (ie alphabet-order: k,j,i,h)
c. Uses number pattern but not looking for operation.	c. OK.	c. OK.	c. "Depends on what k is", or writes $k \longrightarrow x$.
d. OK			d. Just writes -3. e. OK.

Question 5

A rough method of converting from degrees Centigrade to degrees Fahrenheit is to multiply by 2 and add 30.
Can you write this as a formula?

a. Able to work numerically only.

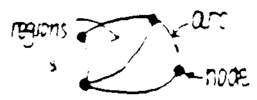
b. Derives a formula but either ignores some information, or otherwise inadequate.

c. Formula wrong way round, eg $C = 2F + 30$.

d. OK, eg $F = 2C + 30$.

Question 6

The formula relating the number of regions (r), the number of arcs (a) and the number of nodes (n) is $r = a - n + 2$.



i. If a=8 and n=5, what is r ?	iii. If there are 6 regions and 4 nodes how many arcs must there be?	iv. Re-arrange the formula... Write it as a = ..	v. If you have a network and you add another node and two more arcs, how many extra regions do you get?
a. Use diagram only.	a. Uses diagram only.	a. Swops letters about.	a. Tries single diagram only.
b. Uses diagram and formula.	b. Uses formula, trial and error.	b. Swops letters about, tries to get an arrangement that works.	b. Tries several diagrams.
c. Uses formula only.	c. Uses formula, inspection.	c. Uses numerical examples, then maps numbers onto letters.	c. Tries to use formula but regards a as 'arcs', not number of arcs, etc. Hence changes a to 2a, n to 3n.
	d. Uses formula, transposes, but unsuccessfully.	d. Transposes, using algoritnm.	d. Uses numerical values in formula successfully.
	e. Transposes successfully.	e. Transposes, seems to understand transformations.	e. Uses formula without giving a and n numerical values, ie $\delta r = 2 - 1$.

Question 8		Question 9	
What is the relationship between x and y if		When are the following true?	
i. $x + 5 = b$ $y + 5 = b$	ii. $x = a$ $y = b$ $x+a+y+b = 12$	i. $a+b+c = c+a+b$	ii. $m+n+q = m+p+q$
a. No relationship.	a. Ignores part of information.	a. Uncertain whether a the same on both sides.	a. Not true -the letters/numbers are different.
b. Tries one value for b.	b. Tries one numerical example, but does not allow x and y, or a and b to be the same.	b. Tries one set of numerical values only.	b. True, when n=p.
c. Tries several values for b.	c. Tries one numerical example (usually all 3's)	c. Tries several sets of values.	c. True, when n=p for negative values also.
d. Solves by matching.	d. As c, but concludes x=y.	d. "Same letters".	
	e. Lists several numerical solutions but no relationship	e. True also for negatives.	
	f. Lists numerical solutions and finds relationship.		
	g. Substitutes algebraically, but unhappy about relationship.		
	h. Substitutes algebraically, finds and accepts relationship.		

An indication of how successfully the responses have been ordered and how the responses to different items relate to each other, is given by the first of the two diagrams below, which shows the category of response given by pupils to 6 items from Draft 1 (1 item from each of Questions 2, 3, 5, 6, 8 and 9). Though quite a few "cross-overs" exist, most are not too severe. The second diagram provides a simplified version of the response-pattern.

Fig 2.4 Children's responses to 6 items from Draft 1.

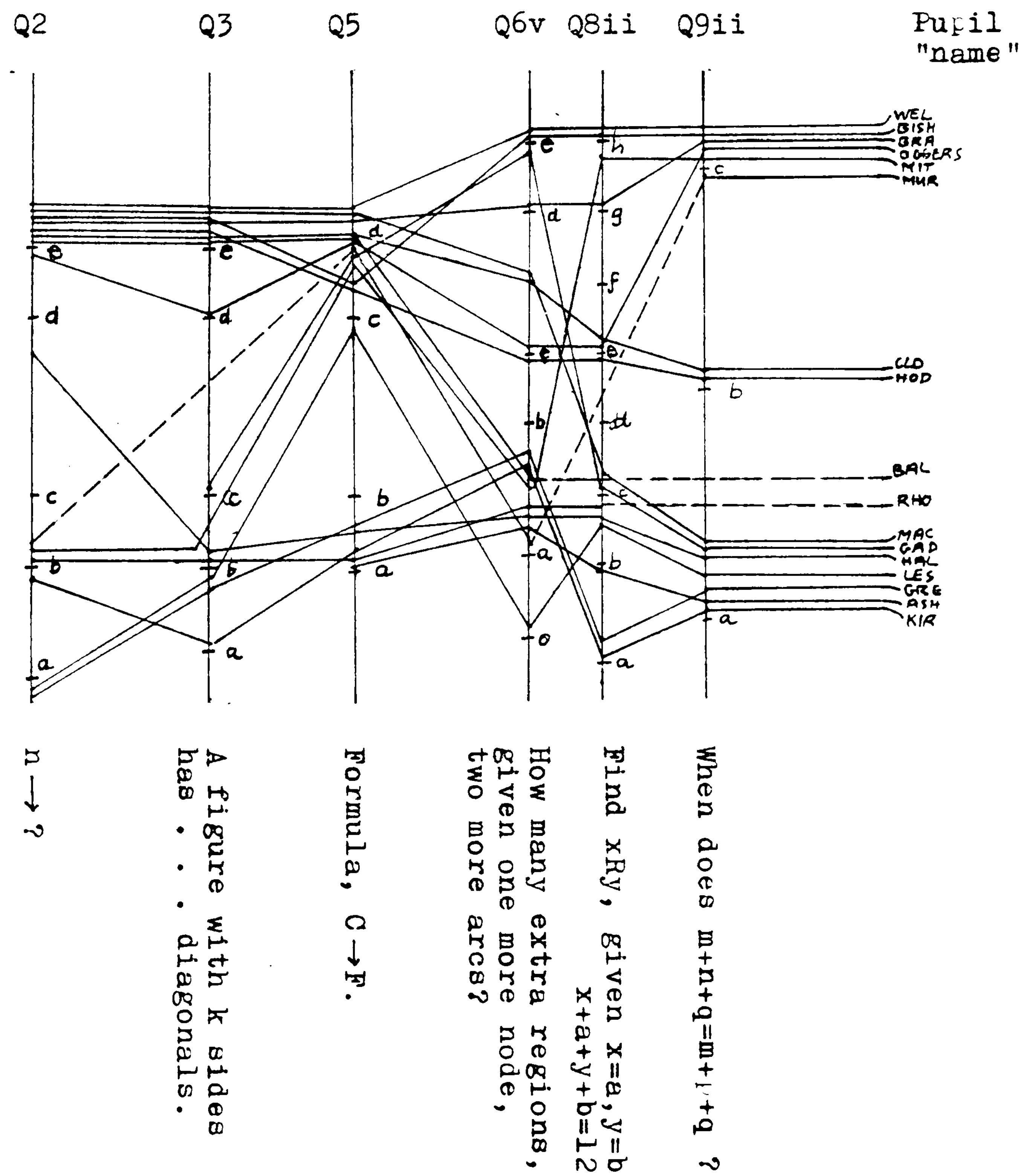
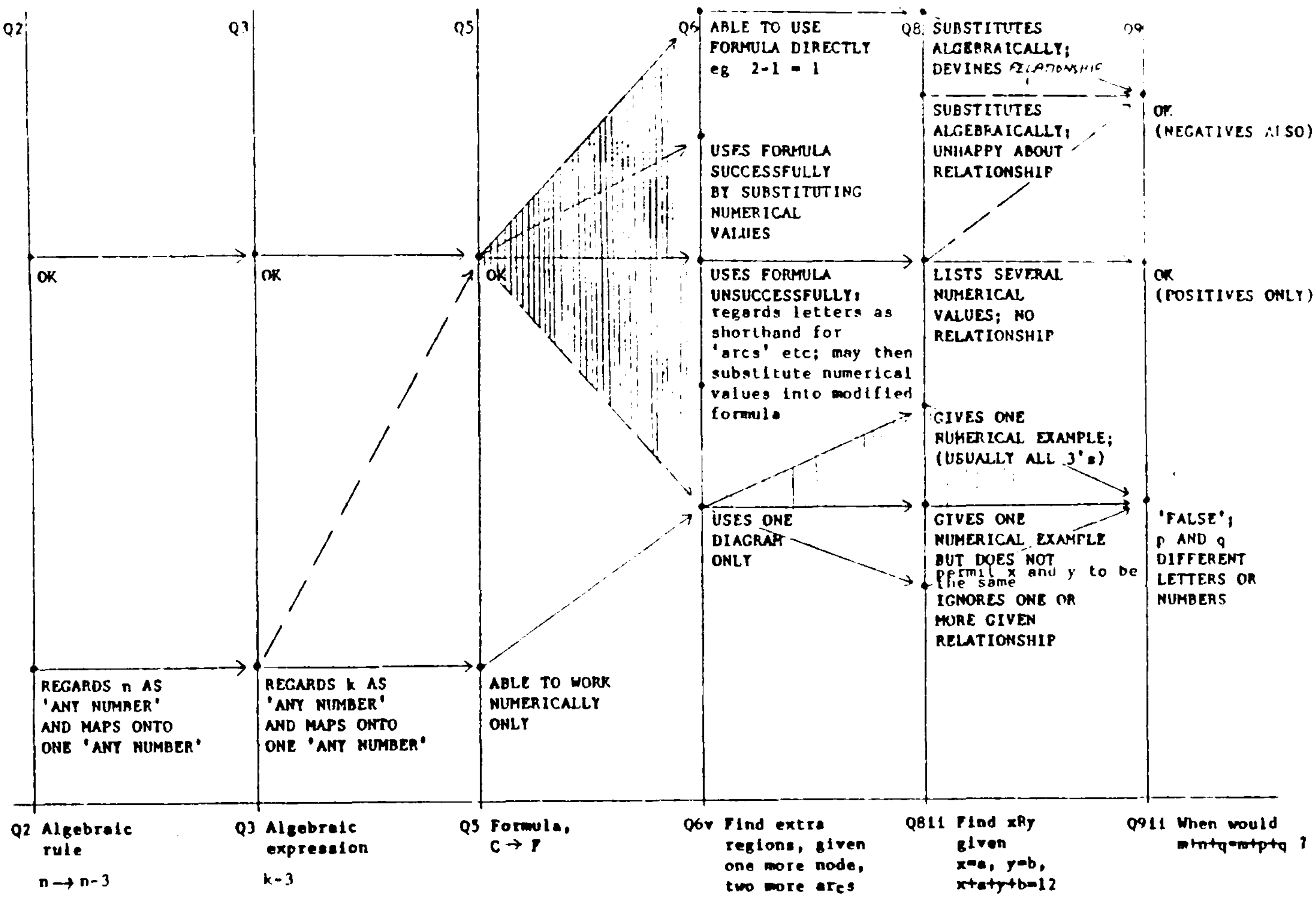


Fig 2.5 Simplified response pattern to 6 items from Draft 1.



Perhaps the most interesting question used in the interviews is Question 6, particularly the last part (If you had a network and you add another node and two more arcs, how many extra regions do you get?). Here is the protocol of part of an interview with a 4th year (15 year old) boy, HOD, who switches between using the letters a and n as numbers (which is what they are supposed to be) and using them simply as a shorthand for the terms "arc" and "node", ie as "objects".

> (Long pause. HOD adds a node and two arcs to the original diagram, and counts the regions.)
> "4 regions.. ?"
> What's the 4?
> "How many extra regions.. oh! extra regions! Just the one."
> Could we use the formula? (ie r = a - n + 2)
> (Pause)
> Say we had.. start with this one again (part iii.). This is a figure with 8 arcs, 4 nodes. Say we did that to the formula..
> "Could be.. r = a - n + 2 + n + a + a (writes)."
> Yes.. that might be the formula you get.. but say we worked it out with numbers.. When we had 8 and 4 (writes r=a-n+2)

then we said r was 6. Now, say we're told to add another node..
"Another node would be 4.. "
How?..
"It says a node equals 4.. and then 2 arcs -it says here an arc equals 8, so that would be 16.. "
(HOD writes $r=a-a+a-n+n+2$, which is an attempt at collecting together the a's etc in his earlier formula.)
(Using $a=8$, $n=4$, he gets $r=10$.)
"Equals 10."
That doesn't give us our 1 does it.. ?
"We won't be able to do it.. "

Another boy who uses the letters as "objects" is OGGERS. As with HOD, he is able to determine that there would be an extra region by drawing a diagram. He is then asked

Could we have used the formula?
(OGGERS writes $r = a - n + 2$, and then $r = 3a - 2n + 2$.)
"-another one node, so that makes 2 nodes, 2n; two more arcs, 3a."

OGGERS then manages to retrieve the situation, but subsequently gets confused between regions and extra regions:

I'm not clear on that.. ($r = 3a - 2n + 2$).
"You get the 'a' number of arcs.. add another 2, so it would be 'a+2'.. That (3a) would be wrong, cos I'd be timesing, 3 times a.. "
(OGGERS writes $r = a+2 - n+1 + 2$, rather than $r=a+2-n-1+2$.)
OK. How does that help us find out how many extra regions?
"You put a number in there.. any number.. "
(Uses $a=3$, $n=1$, in $r=a+2-n+1+2$, gets $r=7$.)
"7 regions.. no, you can't! Can't use, cos that just finds how many regions there are. You can't use this, it just brings you out to 'r'."

The use here of letters as objects had not been anticipated: the item had been constructed in the light of Collis's description of letters used as "variables" (as discussed on page20 in relation to the formula V=LxBxH). One boy who comes close to using the letters in the way Collis describes is BISH:

(Writes 1.)
?
"I substituted examples."
?
"I used that formula using those examples (the earlier

ones), adding 1 and 2 respectively.. And I tried it with another one to see if it was constant."
OK for all then?
"Yeah I think so, because what you're effectively doing, you're adding 2 and 1 -you're adding 2 to the a-n.. -it's really 2-1 so you're adding 1."

Having discovered that some children were prone to use letters as objects, it was decided to write a question that examined this tendency more critically. Thus Draft 3 contained the following question, which also appeared on the final version of the test (as Question 10).

Fig 2.6 Question 10.

Cabbages cost 8 pence each and turnips cost 6 pence each.

If c stands for the *number* of cabbages bought
and t stands for the *number* of turnips bought,
what does
8c + 6t stand for? ....

What is the total number of vegetables bought? ....

As predicted, the tendency to use the letters as objects in this question was exceptionally strong (it might be argued too strong, since many children who performed at a very high level on the final written test still fell into the trap that the question posed). A typical response was that of HAR, who like HOD switched unblinkingly from using the letters as objects to to giving them a numerical value:

(HAR writes 8x8 + 6x6.)
Why is c 8?
"c stands for cabbages and cabbages equals 8 pence.. "

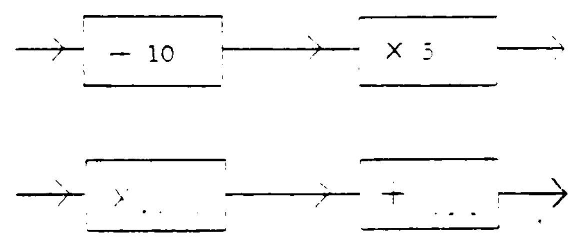
Other children, for example LEW, simply interpreted the expression 8c + 6t as "8 cabbages and 6 turnips".

Another difference between Draft 3 and the first draft of the Algebra test was the question shown below, which was constructed in the light of Collis's statement that a child at the stage of "formal operations" (Piaget's substage 3B) "is able to work on the operations themselves and does not need to relate either the elements or the operations to a physical reality" (Collis,1974,page 6). This question also appeared on the final version of the test (as Question 23).

Fig 2.7 Question 23.

You can feed any number into this machine:

Can you find another machine that has the same overall effect?



Again the question proved to be very difficult, with many children simply switching the operations +10 and x5 around for the second machine. Such a response is perhaps not unexpected though it did come as a surprise that few children seemed concerned about the consistency of their answer, which could have been tested by feeding a number (the same number!) into both machines.

The other major difference between Drafts 1 and 3 was that the later draft included a question on algebraic notation (given g=3 and h=5, children were asked to evaluate the expressions g+h, 2g, gh, h^2 and g(g+h)). Due to pressures of space, and the feeling that the mere knowledge of such notation was not very interesting, the question was later abandoned, although a shorter version was eventually used as a trial item to remind children, in particular, of the convention used for multiplication (eg that 4a means 4 x a).

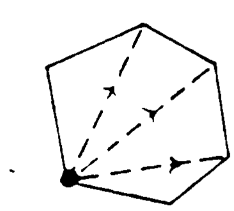
The Class Trials

After the interviews the test was given to whole classes at a time, with children being asked to respond by writing their answers on the test paper. During this process the test went through another 7 versions (Drafts 4, 4.1, 4.2, 4.3, 4.4, 4.5 and 10), one of which (Draft 4.3) is shown in Appendix 2.2 . Draft 4 was very similar to Draft 3 that had been used in interviews, but gradually the characteristics of the test changed. In particular it was decided to reduce the length of questions and at the same time to increase their number, as this would allow a greater variety of items to be used. Also, within each question an attempt was made to reduce or remove the dependence of any one item on its predecessors so that an error made at the beginning of a question was less likely to be carried through to the remaining items. One consequence of these moves was that there was less opportunity to set up interesting contexts in which to place the items; given that it is difficult anyway, for the age-groups under consideration, to find compelling reasons for using algebra, this was a pity.

An illustration of how the questions were reduced is the example shown below, which is the version of Question 3 (on Drafts 1, 2, 3 and 4) as it appeared on Draft 4.1. (The question was extended again, but only slightly, for the final version of the test where it appeared as Question 15.)

Fig 2.5 Question 15 as it appeared on Draft 4.1.

In a shape like this you can get the number of diagonals by taking 3 away from the number of sides.



How many diagonals can you draw from a corner if a shape has k sides? ....

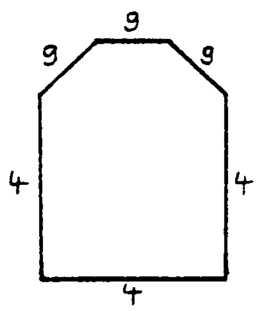
Another long question, Question 6 (concerning the relationship between the number of arcs, nodes and regions), was changed completely. An obstacle to using the question as part of a class test is the difficulty of determining whether answers are being derived diagrammatically, numerically or algebraically. With respect to 6v in particular, various attempts were made to remove this ambiguity while preserving the essence of the item. One such attempt is shown below, which first appeared on Draft 4.3:

Fig 2.9 A version of item 6v Draft 1, as it appeared on Draft 4.3.

For the perimeter of this shape we can write $p = 3g + 12$.

What happens to p if g is increased by 2?

....



Unfortunately, the method of solution seems to be just as ambiguous and it was eventually decided to use the question shown below (Question 19 on the final version of the test) which had also first appeared, in a slightly different form, on Draft 4.3. However, this question too suffered from certain shortcomings as was revealed when the final version of the test was marked and fully analysed.

Fig 2.10 Two more versions of item 6v Draft 1 (Question 19).

$a = b + 3$. What happens to a if b is increased by 2? ....

$f = 3g + 1$. What happens to f if g is increased by 2? ....

Some of the questions were completely new, such as the one below, which first appeared on Draft 4.1 and which was designed to test, in as simple a way as possible, Collis's notion of ALC with respect to operations on unknown numbers. A similar but shorter version appeared on the final version of the test (Question 4).

Fig 2.11 Question 4 as it appeared on Draft -.1.

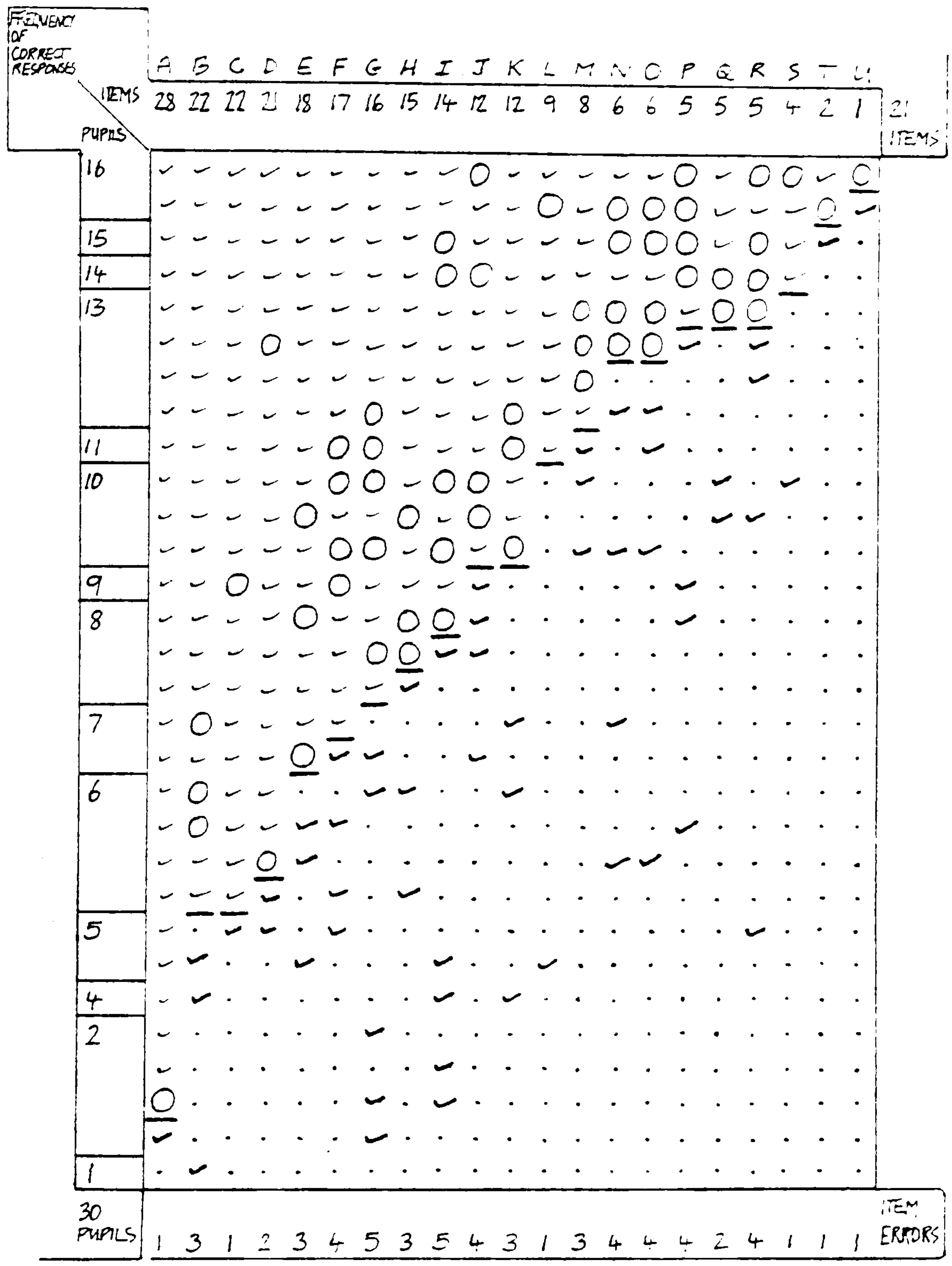
Add 5 onto each of these numbers:

15	4	n	23	n+2	3n	n-6	n+n
.....	.....	.....	.....	.....	.....	.....	.....

In much the same way as the interviews were undertaken to reveal the different strategies that children used (rather than just to determine whether an item was answered correctly, or was easy or difficult) so, when marking the class tests, care was taken to note children's wrong as well as correct responses. The extent to which an item produced common (as opposed to seemingly random) wrong responses, and the extent to which it was possible to interpret these responses in terms of an underlying strategy was one important factor for deciding whether an item should be kept, modified or abandoned. The wrong answers were also examined to see whether the items were being misconstrued (rather than being interpreted at a lower conceptual level) through some unforeseen weakness in the wording.

The items were also examined for "consistency", in the sense of whether the pupils who answered an item correctly were those who performed best on the test as a whole. Consistency was assessed by drawing diagrams like the one below, in which the items and the pupils are ranked in terms of the number of correct responses received and given. (In this example, the items were a subset from Draft 4.4, and the pupils were an above-average class of 13 year olds). The horizontal line partitioning each column on the diagram indicates the position above which all the responses should be correct and below which they should all be incorrect for the corresponding item to have perfect

Fig 2.12 Scalogram for items from Draft 4.4.



consistency. Essentially, the method being used here was a form of Guttman Scalogram Analysis (but without the benefit of a scalogram board or a computer..). The columns were inspected for the incidence of "error" responses (the number of incorrect responses above, or correct responses below, the partitioning line), in terms of which the least consistent items are G and I. (Infact, because the partitioning line for item I separates pupils with tied scores, one of whom answered the item correctly, it can be argued that I is slightly better than G.)

Fig 2.13 Item G from scalogram, and modified version (item 5i).

Item G	Modified Item G (5i on the final version of the test)
a + b = 43	a + b = 43
2a + 2b = ...	a + b + 2 = ...

A modified version of item G was written (see above) which turned out to be far more consistent. However, though numerous scalogram diagrams were drawn for the drafts of the class test, it must be said in retrospect that whole procedure was somewhat limited in value. For a start, the small numbers of pupils (in the above case 30) means that not too much reliance should be placed on the frequency of error responses. Also, simply comparing these frequencies ignores the fact that the number of error responses that an item can have is limited by its facility: for example, item U in the above scalogram diagram, which had one error, can only have one error. No really convincing way of solving this problem could be found, and though several putative solutions have been come across since, none are totally satisfactory (see Chapter 11, where Guttman Scalogram Analysis is discussed in detail).

The TAMS Items

Items from two of the NFER TAMS tests (for 11 and 13 year olds respectively) were analysed at about the same time as the class test version of the Algebra test was being developed. Items were classified into 4 main types and an attempt was then made to find more specific characteristics to account for differences in facility. The analysis resulted in the following table (only some of the items are shown, and these have been modified slightly to protect their confidentiality).

Fig 2.14 Classification of TAMS items.

Approximate Facility (%) 11 year olds 13 year olds	SUBSTITUTION	GENERALISATION	VERIFICATION	TRANSLATION (words → symbols)
A 70,90	ONE whole number for ONE letter ONE simple operation 70,× 2b means 2 × b; b = 3 ⇒ 2b = 6; b = 7 ⇒ 2b = ... 66 × H means HALVING; so H(6) = 12; H(4) = ...			
B 45,70	SEVERAL whole numbers for SEVERAL letters ONE whole number for algebraic EXPRESSION LETTERS directly for numbers ×,69 If a = 1, b = 6, c = 4, a + b − c = ...	Generalisation of pattern to new numerical instance (ONE dimension) 43,× 1→0, 2→1, ... 5→?	Verification of rule by spontaneous substitution of numerical value (ONE sub enough) 46,× ODD+ODD = ODD? EVEN? EITHER? 33,× n+1 > n is TRUE OR FALSE?	
C 10,45	As B, but with notation difficulties (eg quotients, powers, etc) TWO-STAGE substitution ×,43 If a = 3, b = 12, c = 9, $\frac{ab}{c}$ = ... ×,40 y = 2x + □; y = 11, x = 3, □ = ...	As B, but TWO dimensional 15,× 4 * 3 = 6 2 * 5 = 5 4 * 5 = ...		DIRECT trnsltn of arithmetic rule or expression 09,43 1 more than 5 is 5+1 1 more than X is ...... 03,45 1→0, 2→ x → ? ×,58 I have x pence you have y pence; together we have ......
D 0,<30			–by spontaneous but SYSTEMATIC substitution ×,29 Which is ODD if n is ODD OR EVEN: 3n, n+2, n^2, 2n+1, 6n+1	As C, but operations on unknowns not made explicit ×,29 Tea. one spoon per person and one for the pot How many spoons (M) if n people?
E 0,<10	ONE EXPRESSION for another, *or* Matching of TWO expressions and then substitution of numbers for letters in another exprssn ×,0? If [illegible]			

The outcome of the analysis was seen as encouraging, in that the characteristics that had been identified seemed to fit items that were quite neatly clustered in terms of facility; also, <u>within</u> each of the 4 main item types the characteristics were related to each other in a quite coherent way (eg the Substitution clusters formed a natural progression, from items involving a single letter, to several letters, to more complex notation, to the substitution of expressions). However, the relationship <u>across</u> the item types (between the characteristics at each facility level A, B, C, D and E) was far from clear, and in this respect the analysis sounded a note of caution: fragmenting a test into item types of this sort would perhaps not be appropriate for designing the Algebra test since it was difficult to see how these types (and equally important ones like "Equation Solving", "Translation, of symbols into words", etc) could adequately be investigated by a single test. Visions of a full-blown classification system reminiscent of Bloom's (Bloom et al,1956) loomed ominously, but fortunately it was beginning to look as though some of the ideas derived from Collis might provide a more manageable but still useful focus.

The Final Version of the Algebra Test

The first draft of the test, written in the middle of 1975, contained several questions that can best be described as exploratory. By the end of the year the rationale of the test had taken definite shape. Specifically, items were being written to explore the different ways in which children interpret the letters in generalised arithmetic. Here are some notes made in December 1975, at about the time Draft 4.3 was written.

Fig 2.15 Children's interpretations of the letters in generalised arithmetic (notes made in December 1975).

Possible LEVELS of INTERPRETATION of PRONUMERALS

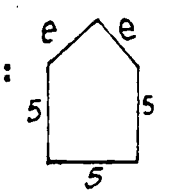
A letter → UNIQUE number which has to be DETERMINED
eg Add 5 to n → 19

B letter → IGNORED, expressions manipulated by ASSOCIATION
eg Add 5 to 3n → 8n

C_1 letter → like APPLES, etc
eg 3a + 1b → 3 apples + 1 banana
3a + 1a → 4 apples

$C_{1'}$ letter → OBJECT per se
eg 3a + 1b → 3 a's and 1 b

C_2 letter → NAME or QUALITY
eg perimeter (p=15+2e) of this shape:
e thought of not as quantified length
but simply as name for side.



C_3 letter → SUMMARISING NAME for numbers
eg 1 → 3   x seen as "the numbers" (or "a number")
2 → 4   on the left hand side.
5 → 7
x →

C_4 letter → SUMMARISING NAME for number/quality
eg sides → diagonals
10 → 7
57 → 54
s →

C_5 letter → SPECIFIC UNKNOWN number
eg Add 5 to n → n+5

C_6 letter → SPECIFIC UNKNOWN number/quality
eg polygon with s sides has s-3 diagonals from one vertex.

C_7 letter → GENERAL UNKNOWN
eg a=b+3 → a,b could be 4,1 5,2 6,3 etc.

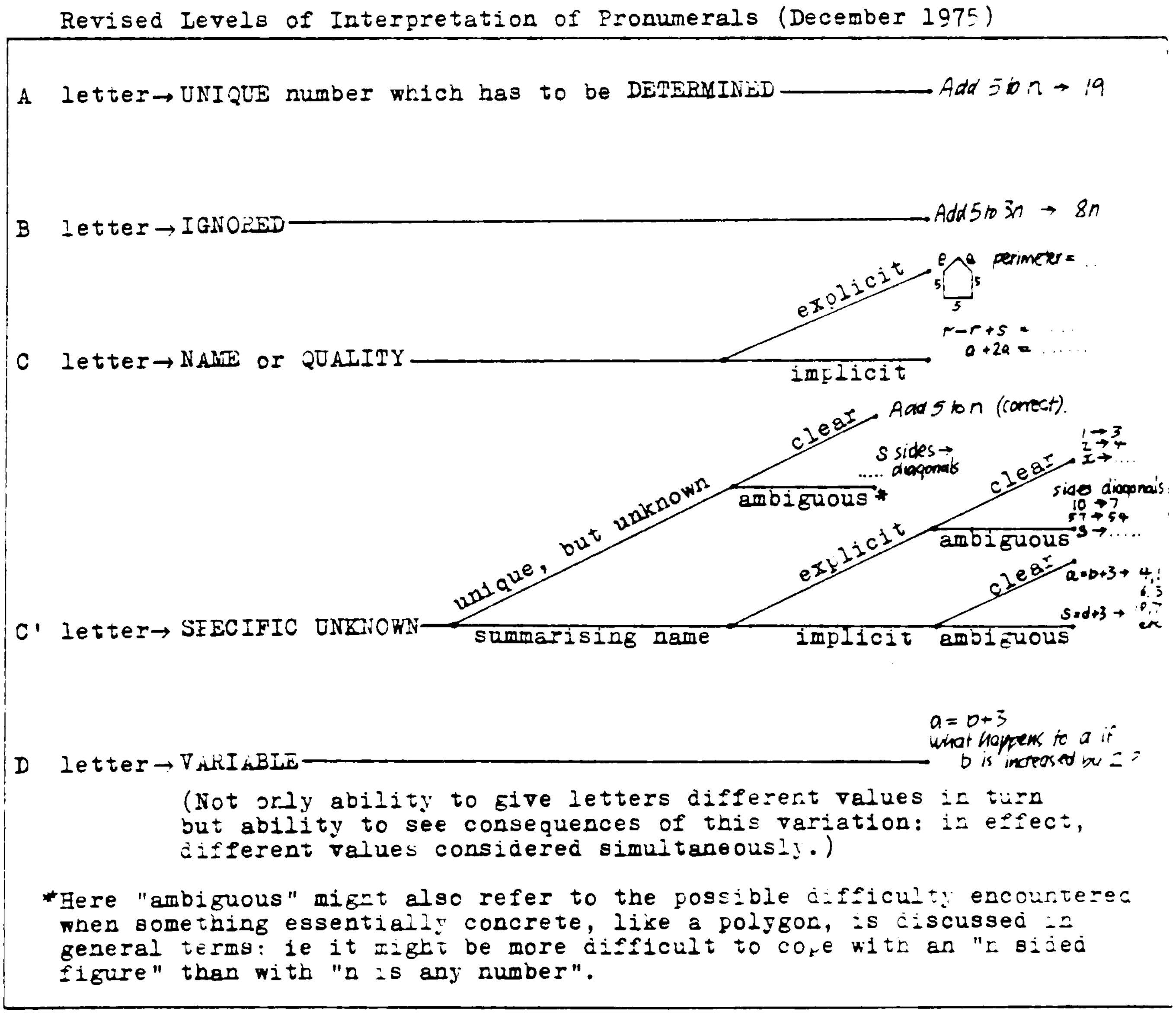
D letter → VARIABLE
eg a=b+3, b increased by 2 → a increased by 2.

The notes continued with a discussion of the similarities and important differences between the levels. The C levels were split into two major groups: C_1, $C_{1'}$ and C_2 and C_3 to C_7, the latter being seen as representing the first genuine use of letters as unknown numbers. However, it was felt that there was still an important distinction between this and the use of letters as variables (level D). For the item

a = b + 3. What happens to a if b is increased by 2 ?
it was argued that at Level D the child is aware that a and b can have many values and that the relationship between them remains the same (a 2nd-order relationship is divined between a and b, albeit a simple one): "a is ALWAYS 3 bigger than b", so if b increases, a increases by the same amount. On the other hand, at level C_{3-7} a and b are unknown but essentially fixed numbers, so the relationship is simply seen as a relationship between two (unknown) numbers: "THIS a is 3 bigger than THIS b", so if b changes, the relationship is destroyed.

The notes concluded with the following diagram:

Fig 2.16 Summary of children's interpretations (December 1975).



At a later date some of the levels were re-named: level A became Letter EVALUATED, B was called Letter NOT USED, and C was called Letter as OBJECT. It was also decided to revive level C_7, which was re-named Letter as GENERALISED NUMBER and referred to the notion that a letter could take several numerical values (though without necessarily being regarded as representing such values "simultaneously"). Though these different ways of interpreting the letters could be partially ordered, it was also decided to drop the term "level" since this falsely conveyed the impression that the interpretations formed a strict hierarchy. Instead, the rather unwieldy term "category" was adopted, whilst "level" was reserved for items of similar cognitive demand (in which the letters might be interpreted in a number of different ways). Thus eventually these six categories were arrived at:

Letter EVALUATED
Letter NOT USED
Letter as OBJECT
Letter as SPECIFIC UNKNOWN
Letter as GENERALISED NUMBER
Letter as VARIABLE

The final version of the Algebra test contained items in each category, in the sense that the category described how the letters had to be interpreted for the item to be answered correctly and/or described common wrong answers. For example, the first two items of Question 5 on the final version could be solved by not using the letters (Collis's matching strategy), whilst the third involved operating with at least a specific unknown (g added to 8); moreover, on this item children commonly evaluated g (giving answers like 12, 15 and 9 instead of 8+g).

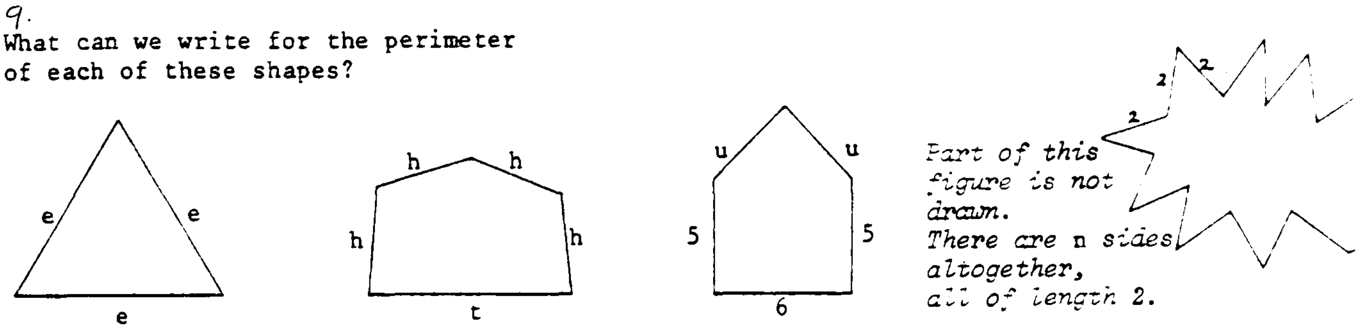
Fig 2.17 Question 5.

If $a + b = 43$	If $n - 246 = 762$	If $e + f = 8$
$a + b + 2 =$	$n - 247 =$	$e + f + g =$ ......

In the first three items of Question 9 the letters can be interpreted as objects, but in the last item the letter is again a specific unknown. In Question 11 children are asked to evaluate the letters (which can be done without operating on the letters as unknowns), whilst Question 16 tests the notion of generalised number and Question 3 the notion of variables.

Fig 2.18 Questions 9, 11, 16 and 3.

9.
What can we write for the perimeter of each of these shapes?



p = ...   p = ...   p = ...   p = ...

11.
What can you say about u if $u = v + 3$ and $v = 1$ ....

What can you say about m if $m = 3n + 1$ and $n = 4$ ....

16.
What can you say about c if $c + d = 10$ and c is less than d ....

3.
Which is the larger, $2n$ or $n + 2$? ....

Explain: ....

A complete copy of the final version of the test is shown in Appendix 2.3, and the test and the categories themselves are discussed fully in Chapters 4 and 5.

Chapter 3 ADMINISTRATION OF THE ALGEBRA TEST AND THE SAMPLES USED IN THE LARGE-SCALE SURVEY

The final version of the Algebra test was given to large samples of children in their 2nd, 3rd and 4th year of secondary school. Most of the testing occurred in the first half of the summer term 1976, with a few children taking the test at the end of the spring term. Three other CSMS mathematics tests (Graphs, Ratio, Vectors) had been developed by this time (in addition to the Numbers 1 test that came out the previous year), and most of the children taking the Algebra test also took at least one of these. (A few children also took a number of Piagetian class tasks as part of a survey being undertaken by the science wing of CSMS, and comparisons were obtained with subsequent CSMS mathematics tests by using the Algebra test again in 1977, though on a smaller scale. This second testing also provided some longitudinal data).

Most of the schools that took part in the large survey in 1976 were found through members of the CSMS mathematics team meeting teachers at in-service courses, etc. Schools were selected from different parts of England, and there was a mix of rural, urban and city schools, though no rigorous attempt was made to control for factors like socio-economic background, type of school, curriculum, or sex. With the exception of a selective and a super-selective school, London schools were excluded.

It was decided to assess the representativeness of the samples in terms of standardised scores on a non-verbal reasoning test, and if necessary to adjust the samples so that the frequency of scores did not differ significantly from what would be expected for a representative sample. Unfortunately it was not possible to find a test suitable for all the year-groups being tested. A compromise was reached by choosing the test already being used by the science wing of CSMS (the Calvert DH test published by the NFER; Calvert,1958) and giving this to all 2nd year children in every school in the survey, on the assumption that the score distribution of all the 3rd and 4th year children in the same school would not be significantly different. In support of this assumption schools were asked to confirm that no obvious changes had befallen their intake. Also, the schools were asked to ensure that, apart from absentees, all children were given the Algebra test in the year groups that the schools had offered to test, even though this might include children for whom the test would be inappropriate.

Nearly all the participating schools were visited by at least one member of the CSMS mathematics team in order to explain the purpose of the testing and how the test should be administered. Where possible testing occurred during a normal mathematics period and was administered by the children's own teacher who was provided with the following notes:

> The test, including the two trial items (on a separate sheet), should take about 40 minutes. This time need not be adhered to too precisely though it is not desirable to have most of the class, having done all they can, sitting restlessly doing nothing...
>
> Please make it clear to the pupils that this is not a test in the usual sense -we are trying to find out how each individual pupil interprets the items, rather than simply mark him right or wrong. Many of the items may appear unfamiliar to the pupils; however, encourage them to put down what they think might be a sensible answer even if they are very unsure of what the 'right' answer might be. Discourage pupils from leaving a lot of blanks.
>
> Some younger pupils may find the test rather strange, difficult and perhaps disconcerting; explain to them that this is a test also being done by much older children, so we don't expect them to know all the answers -nonetheless we are interested to know what they think .

Equally some older pupils might feel afronted by some of the simpler items; again explain that this is a test designed for pupils of widely differing ages and "there have to be some easy questions for the younger pupils". Encourage these pupils to take the test seriously as there will almost certainly be some items at which they won't be entirely successful.

The two trial items should be given to all classes. With the older children only 2 or 3 minutes need be spent on them, with the younger classes perhaps 8 or 10 minutes may be required. After the whole test is completed the trial item sheet can be thrown away.

Trial item 1 is intended to remind pupils of the convention that 4a is to mean "4 times a". Also many younger pupils think that a letter stands for a unique (albeit unknown) number; eg a=1, b=2, c=3, etc. Hopefully the trial item will make clear that this need not be so. When the item has been attempted discuss it briefly and give and explain the answers.

The most important aspect of trial item 2 is the mapping "$n \rightarrow n+3$". Many younger pupils will be bemused by this and will put answers such as 3 (ignoring n) or q (counting on 3 letters) or 17 (n is the 14th letter of the alphabet) or simply any number. Explain that we don't know what n is so the best we can do (however unsatisfactory it might appear..) is to write n+3.
Encourage pupils to work through the trial items as quickly as they can, and not to worry about the answers too much -they will be explained to them short and their answers will not be marked.

The two tables below show the composition of the Algebra samples and briefly describe the nature of each school (the schools are coded to preserve confidentiality). All three sampl include a class from a super-selective boys' secondary school (school 23) that was added after it had been found that the standardised Calvert test score distribution for each of the corresponding 2nd year samples was underrepresented by scores at the top end of the scale. After this adjustment it was found in each case that the distribution of standardised scores was r significantly different from that expected for a representative sample of children ($p > .20$, .15 and .20 for the samples corresponding to the 2nd, 3rd and 4th year Algebra samples respectively, using the Kolmogorov-Smirnov one sample test of goodness of fit; Siegel,1956).

Fig 3.1 Composition of 1976 2nd, 3rd and 4th year Algebra samples.

Algebra Samples 1976

2nd YEAR SAMPLE		3rd YEAR SAMPLE		4th YEAR SAMPLE	
School Code	Number of Children	School Code	Number of Children	School Code	Number of Children
05	182	04	140	01	21
06	220	05	160	06	92
07	82	06	202	09	98
08	123	08	133	11	134
15	86	17	92	14	142
16	276	18	212	16	220
20	133	23	22	23	24
23	26				
Total	1128	Total	961	Total	731
Standardised Calvert Score of Corresponding 2nd Year Sample					
Mean	SD	Mean	SD	Mean	SD
99.5	14.1	100.2	14.8	99.3	13.8

Fig 3.2 Nature of schools in 1976 Algebra sample.

School Code	Type of School and Region	Standardised Calvert Score of 2nd year Mean	SD
01	Mixed Comprehensive (Bristol)	100.2	13.4
04	Mixed Comprehensive (Herts)	95.1	13.4
05	Boys' Comprehensive (Herts)	93.2	13.7
06	Mixed Comprehensive (Herts)	100.3	13.3
07	Boys' Comprehensive (Herts)	100.0	
08	Mixed Comprehensive (Herts)	97.1	12.2
09	Mixed Comprehensive (Herts)	101.1	12.8
11	Mixed Comprehensive (Herts)	98.2	13.3
14	Mixed Comprehensive (Glos)	96.5	13.2
15	Girls' Selective (S.London)	108.8	
16	Mixed Comprehensive (Covntry)	98.8	13.3
17	Girls' Selective (Devon)	116.9	
18	Mixed Comprehensive (Notts)	101.0	13.8
20	Mixed Middle (Leeds)	99.7	14.9
23	Boys' Super-Selective (London)	122.8	

Note: Calvert standardised to a mean of 100, SD 15.

Chapter 4 THE MARKING SCHEME FOR THE ALGEBRA TEST

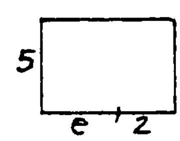
Sixth form mathematics students and undergraduates were employed through most of the summer vacation 1976 to help mark the Algebra scripts used in the large-scale survey (and the scripts of the three other CSMS mathematics tests put out at that time). This meant, amoungst other things, that each day at least one member of the CSMS mathematics team had to be in the office by 9 o'clock, armed with coding sheets, correction fluid and orange squash.

The Algebra marking scheme developed alongside the test and went through about the same number of drafts. Before the scheme was finalised several hundred completed scripts were examined for unexpected answers, and where these answers made sense an attempt was made to incorporate them into the marking scheme. It took several days to train the students to use the scheme properly and subsequently one in ten scripts was checked to ensure that the scheme was still being used correctly. After sufficient practice a script could be marked in about 3 minutes. More recently a simplified version of the scheme has been produced for the NFER (Küchemann, in press), which it is hoped can be used without extensive training. The original scheme is shown in Appendix 4.1 .

The scripts were marked straight onto computer coding sheets and the digits 0 to 9 were used to code the answers to each item. Correct answers were coded 1 (and sometimes also 2); 3 was used for answers that were ambiguous or in some way inadequate without being explicitly wrong; 9 was used for miscellaneous wrong answers and 0 when no answer was given at all. The other digits were used in a variety of ways, though commonly they were used as follows: 4 for incorrect punctuation (eg 3nx4 instead of 3n+4 for the item "Add 4 to 3n"); 5 where numbers were combined without taking proper note of the letters (eg 7n instead of 3n+4 6 where the letter was ignored entirely (eg 7 instead of 3n+4); 7 where letters were evaluated or transformed according to their position in the alphabet, and 8 for other, wrong numerical answers. The scheme used for item 7iv conforms to this pattern in most respects and is shown below:

Fig 4.1 Marking scheme used for item 7iv.

What is the area of this shape?



Code	Response	Meaning of Code
1	5(e+2) 5e +10	Correct answer
2		(Where appropriate, Code 2 is given for "weak" correct answers.)
3	5e+2 e+2 x5 5e2 MISUNDERSTOOD: PERI.	The answers are ambiguous due to inadequate punctuation: the child may have been trying correctly to express e AND 2 multiplied by 5.
4	5e x 2	Here the punctuation is not so much inadequate as definitely wrong.
5	e +10 10e 7e	For e+10 the numbers have been combined correctly but the letter has essentially been ignored. The same thing seems to be happening to the letter in the other answers though they occurred less often.
6	10 7	The numbers have been combined but the letter has been ignored entirely.
7	35 50	e seems to have been given the value 5 from its position in the alphabet.
8	25 20 30	e seems to have been evaluated geometrically (the base of the rectangle looks to be about 5 units long).
9		All other wrong answers
0		Item omitted

Note: responses written small occurred less frequently but shared some of the characteristics of the other responses under the same code.

A few items were marked in a much simpler way. For example the first two parts of Question 7, whose sole purpose was to check that children knew how to work out the area of specific rectangles (3 by 4 and 6 by 10), were just marked as correct (code 1), wrong (code 8 rather than code 9, as the answers are numerical) or omitted (code 0). On the other hand, the scheme for some items, such as 17i, was much more complicated:

Fig 4.2 Marking scheme used for item 17i.

Mary's basic wage is £20 per week.
She is also paid another £2 for each hour of overtime that she works.

If h stands for the number of hours of overtime that she works, and
if W stands for her total wage (in £'s)
write down an equation connecting W and h: ....

Code	Response	Meaning of Code
1	W = 20 + 2h w = 20 + 2×h	Correct answer
2	TWO NUMERICAL PAIRS OF VALUES (CORRECT) W=28 h=4 W=30 h=5, etc OR 28,4 30,5 etc	Though this would seem to be a much lower level answer than that of code 1, it does express the important notion that a letter can take more than one value. (In the event less than 1% of all children gave a code 2 response.)
3	W + 2h = wages W+2h 20+2h EXPRESSIONS INVOLVING Tw, etc (CORRECT). AMBIGUOUS PUNCTUATION: W=20+h2	The answers are ambiguous or inadequate (eg W is used for the basic wage) but the letters are at least being used as numbers (not objects as below).
4	W=20+h 20+h W = wages+h Total wage = W+h W= w+h h→w	An attempt is being made to establish a general relationship but there is a tendency to use the letters as objects: h seems to stand for the hours of overtime rather than the number of hours.
5	W+h. w∩h wh	The answer is again general, but the letters are being combined in the most primitive way (simply "juxtaposed" or "associated").
6	ONE NUMERICAL PAIR OF VALUES (CORRECT) W=28, h=4 etc OR 28,4 etc	The letters have been evaluated, and only one pair of (correct) values is given.
7	2h = 24 W 4h = 28 W etc. 24 = W+2h 28 = W+4h etc.	The letters have again been evaluated (correctly) but the letters are then used as objects: 2 hours overtime gives a total Wage of 24.
8	W=20 h=2 20W+2h W=10h 22 = w+h £22	In some way these answers all involve the given (ie most obvious) numbers, 2 and 20.
9		All other answers, including arithmetical errors.
0		No response.

There is some similarity between the meaning of codes 4, 5, 6, 7 and 8 for item 17i (and 10i, 16, 20 and 22) and the meaning they are more commonly given, as in the case of item 7iv (and 4i, 4ii, 4iii, 4iv, 5iii, 9i, 9ii, 9iii, 9iv, 12, and 15ii). For example the responses under codes 6, 7 and 8 are essentially numerical, whilst for code 5 the elements are combined in a very direct and primitive way (the elements are letters in the case of 17i, and numbers for 7iv but with the letter remaining in the answer).

One of the most important items on the test was Question 3, where children were asked to choose the larger of 2n and n+2 and to explain their answer. The most common response was to select 2n, often with an explanation like "Because it's multiply", and this was coded 4, whilst the other definite choice, n+2 (and the occasional claim that the expressions were "The same") was coded 8. The point of the item was to see whethe children recognised that the relative size of the expressions depended on the value of n. Some children were successful in this but the quality of their explanations differed enormously and were difficult to classify: some just wrote "Depends" or "Usually 2n" with no further explanation, whilst others provided examples involving rather "disjointed" values of n (such as "It depends: eg 2n is larger than n+2 when n=5 and smaller when n=1") which gave the impression that their appreciation of the consequences of n varying was not fully systematic. The critical value of n is 2, at which point 2n and n+2 are equal, and it was eventually decided to classify answers under code 1 only if the explanations referred to several values at or below this critical point (or if the references to n were obviously systematic, as in "2n is the larger when n is greater than 2"). Conditional answers referring to only one value at or below n=2 were classified under code 2, and the remaining conditional answers under code 3. Thus the marking scheme for Question 3 was as follows:

Fig 4.3 Marking scheme used for Question 3.

Which is the larger, 2n or n + 2 ? ....

Explain: ....

Code 1	Code 2	Code 3	Code 4	Code 8	Code 9/0
"DEPENDS" + AWARENESS OF AT LEAST TWO OF n=2 n=1 n=0 n NEGATIVE ("n > 2" OK!)	"DEPENDS" + ONLY ONE OF n=2 n=1 n=0 n NEGATIVE	"DEPENDS" OR "USUALLY 2n" BUT NONE OF n=2, 1, 0, NEG	2n	n+2 "SAME"	OTHERS/OMITTED

The facilities of the items on the Algebra test were usually determined by combining codes 1 and 2. The only exceptions were the items listed below (apart from those items where code 2 had not been used at all):

Fig 4.4 Codes used to determine facility of Algebra items.

Items whose facilities were determined other than by combining codes 1 and 2

Item	Codes used to determine facility
14	1, 2, 3
19i	1
19ii	1
20	1, 2, 3

The facility of each item, and the frequency of each code, for the 1976 2nd, 3rd and 4th year samples (and also for a nonrepresentative sample of 103 5th years) is shown in Appendix 4.2.

Chapter 5 THE INTERPRETATION OF THE RESULTS OF THE LARGE-SCALE SURVEY: LEVELS OF UNDERSTANDING IN ALGEBRA

As will be recalled from Chapter 2, the Algebra test was designed to examine the different ways in which children interpret the letters in generalised arithmetic. Six categories were listed at the end of that chapter and these subsequently provided the basic framework in terms of which children's performance on the test was analysed. However, the analysis was taken a step further: items from the test that correlated well with each other were partitioned into 4 facility bands, with the cut-off between the bands determined by which interpretation of the letters was sufficient to answer the items correctly, and by a second dimension that is related to the notion of ALC and might be called "structural complexity" (Collis,1975a; Halford,1978) which could be applied successfully to most of the items. Items in these 4 bands were used to define "levels of understanding" and it is the purpose of the present chapter to describe these levels, after first discussing the 6 categories in more detail. The statistical methods of analysis used to select the items, and the arguments that might justify the use of the term "levels", are considered in Part B.

To simplify the discussion the response frequencies quoted in this chapter are for the 1976 3rd year sample only (N=961), unless it is stated otherwise. Changes in performance from one age group to the next are examined in the chapter that follows.

Ways of Interpreting the Letters in Generalised Arithmetic

The 6 categories are as follows:

Letter EVALUATED. This category applied to responses where the letter is assigned a numerical value, as when children decide that the area of the rectangle of dimensions 5 by e+2 (item 7iv) is 35, by using the "alphabet code". The category can also be used to describe items, in those cases where a numerical value is asked for but where it is not necessary to manipulate the letter first, as in "Find a if a+5=8" (item 6i).

Letter NOT USED. Here children ignore the letter, or at best acknowledge its existence without giving it a meaning -as when children are asked to "Add 4 to 3n" (item 4ii) and write 7n or just 7, instead of 3n+4. Certain items can be solved successfully in this way, for example "Add 4 onto n+5 (item 4i), where all that is required is that 4 is added to 5, as long as n is not lost from the answer.

Letter as OBJECT. Here the letter is regarded as a shorthand for an object or as an object in its own right, as when "2a+5a" is thought of as "2 apples and 5 apples" or simply as "2 a's and 5 a's, which makes 7 a's altogether". Some expressions can successfully be simplified in this way but other times this use of the letters is quite inappropriate, for example when the letter is meant to stand for the number of an object, as in "a apples", and not the object itself.

Letter as SPECIFIC UNKNOWN. Here children regard a letter as a unique but unknown number, and can operate upon it directly, as is required to solve "Multiply n+5 by 4". (item 4iii) for example, but where many children just operate on 5, giving the answer n+20.

Letter as GENERALISED NUMBER. The letter is seen as represen ing, or at least as being able to take, several values, as is required in "c+d=10, c < d, c=..."(item 16).

Letter as VARIABLE. The letter is seen as representing a range of unspecified values, and a systematic relationship is seen to exist between two such sets of values. A (minimal example of this occurs with the item "a=b+3. What happens to a when b is increased by 2" (item 19i), when the relationship between a and b is interpreted as "a is always 3 bigger than b", rather than "this particular a is 3 bigger than this particular b" which says nothing about their relationship when b changes.

The categories are now discussed more fully, in the light of the responses given by the 3rd year sample. In this, reference will be made to the levels to which items were assigned, which should give some indication of how the categories and the levels relate, although the levels are not discussed in their own right until the next section. The lowest (easiest) level is level 1.

Letter Evaluated

This is one of three interpretations by which children can avoid operating on letters as unknown numbers, in this case by giving the letters a numerical value. This applies to items 6i, 11i and 11ii in the table below, but not to item 14.

Fig 5.1 Children's responses to items 6i, 11i, 11ii and 14.

6i(Level 1) What can you say about a if a+5=8	%	11i(Level 2) What can you say about u if u=v+3 and v=1	%	11ii(Level 2) What can you say about m if m=3n+1 and n=4	%	14(Level 3) What can you say about r if r=s+t and r+s+t=30	%
a=3	92	u=4	61	m=13	62	r=15 r=30-s-t	35 6
		u=2	14	other values	14	r=10	21

As can be seen from the percentages in the table, 6i (level 1) is answered correctly by nearly all the (3rd year) children (all they need do is recall a familiar number-bond or count on from 5 until they reach 8). Both parts of Question 11 are harder (level 2), but are still answered correctly by a majority of the children. The increase in difficulty is probably mainly due to the fact that the items involve two unknowns rather than one, which makes the first equation in each item "ambiguous", in the sense that they are true for more than one pair of values; however, this ambiguity is resolved as soon as the second equation is reached. The answer u=2 to 11i suggests that some children may have had difficulty with the mathematical language, due to a too hasty scanning of the symbols or because of a false translation of the sort

u = v + 3

u and v add-up-to 3.

This statement has the attraction that u and v, when taken together, are equal to something definite: in the original equation, all that is known is that one unknown is 3 bigger than another..

Item 14 is harder still (level 3). It can be solved by replacing s+t by r in the second equation, but this involves handling a letter as an unknown which puts the item into the specific unknown category (though once the substitution has been made, r can be evaluated from r+r=30). The substantial proportion of children who wrote r=10 seem to have avoided this category by evaluating r directly from the second equation (10+10+10=30).

Letter Not Used

In Question 5 the first two parts, but not 5iii, can both be solved by not using the letters. 5i proved to be very easy (level 1) even though it seems to involve two unknowns. However, nothing need be done to these unknowns: they can be eliminated by a matching strategy which focuses attention on +2 by which the left hand side of the equations differ and which is then applied to 43.

Fig 5.2 Children's responses to Question 5.

5i(Level 1) If a+b = 43, a+b+2= ...	%	5ii If n-246 = 762, n-247 = ...	%	5iii(Level 3) If e+f = 8, e+f+g = ...	%
45	97	761	74	8+g	41
		763	13	15	2
		Other values	8	12	26
				8g	3
				9	6

Item 5ii did not correlate well enough to be assigned to a level; however, it can be solved in the same kind of way, by matching the two equations, although there was also evidence that some children evaluated the letter, which tended to lead to arithmetical errors because of the large numbers. The size of the numbers is one reason why the item was more difficult than 5i, but also the operation (-1) is not given explicitly and is counterintuitive because 247 is greater than 246, some children were persuaded to add 1 to 762 instead of subtracting (see also Shiu,1978).

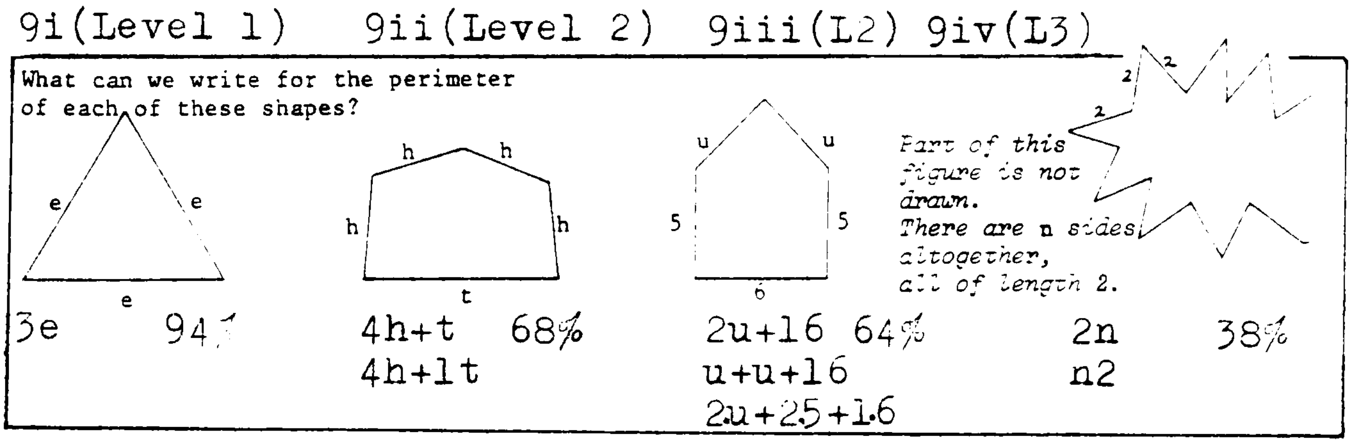
Item 5iii can also be solved by matching, but though e and f can be avoided in this way, children still have to operate with g, which means using a specific unknown. (This puts the item at level 3.) Many children tried to resolve this difficulty by evaluating g, often in a quite logical way but which led to answers like 12 (4+4+4=12) or 15 (g is the 7th letter of the alphabet). Others just added 1, presumably because this is the simplest way of making the answer bigger.

Letter as Object

This category applies to the first three parts of Question 9 (where the letters can be regarded as names for the sides rather than denoting their unknown lengths) but not to 9iv (where the letter is clearly defined as a number).

Fig 5.3 Correct responses to Question 9.

9i(Level 1) 9ii(Level 2) 9iii(L2) 9iv(L3)



The interpretation of the letters as objects also works successfully in some of the parts of Question 13 shown below, in which children were asked to simplify algebraic expressions.

Fig 5.4 Correct responses to parts of Question 13.

13i(Level 1)		13iv(Level 2)		13viii(Level 3)		13v(Level 4)	
2a+5a =	%	2a+5b+a =	%	3a-b+a =	%	(a-b)+b =	%
7a	86	3a+5b	60	4a-b	47	a	23

13i and 13iv can be solved by thinking of the letter(s) as a shorthand for apples (and bananas), say, so that 2a+5b+a becomes 2 apples and 5 bananas and another apple, making 3 apples and 5 bananas in all. Or the letters can be thought of as objects in their own right. This makes the items far easier than if the letters had to be treated as unknown numbers. However, both approaches begin to break down in 13viii (level 3): 3 apples take away 1 banana makes no real sense (unless there already are some bananas..); nor does 3 a's take away one b, unless b is of the same "quality" as a, in other words a number. This difficulty is heightened by the brackets in 13v, which focus attention on an expression which cannot itself be simplified. (However, the brackets are not the major problem: a similar item involving addition throughout, (a+b)+a, was far easier, and quite close in facility to 13iv which is without brackets.)

Using a letter as an object, which amounts to reducing the letter's meaning from something quite abstract to something far more concrete and real, enabled many children to give correct answers to items which they would not have coped with had they had to use the intended meaning of the letter. However, this reduction in meaning often occurred when it was not appropriate. This happened particularly with items that involved objects (cabbages, wages, hours, cakes, pencils) but where it was essential to distinguish between the objects themselves and their number. This distinction can be very difficult to grasp: a classic example is the translation of the relationship "one shilling equals 12 pence" into s=12d (letter as object) instead of d=12s (letter as specific unknown, generalised number of variable). The most difficult question of this sort on the Algebra test was Question 10:

Fig 5.5 Question 10.

Cabbages cost 8 pence each and turnips cost 6 pence each.

If c stands for the *number* of cabbages bought
and t stands for the *number* of turnips bought,
what does

8c + 6t stand for? ....

What is the total number of vegetables bought? ....

Understandably, most children (52%) interpreted the expression 8c + 6t as 8 cabbages and 6 turnips, whilst others (23%) took this a step further and converted it to 100 pence of 1 (8x8 + 6x6). Some answers, like "The cost", were ambiguous so children were also asked for the total number of vegetables bought (c+t). Only 4% gave the correct answer, whilst 73% gave the answer 14.

The same confusion arose in Question 22 (level 4), even with children who performed well on the test as a whole. To solve the question the letters have to be regarded at least as specific unknowns: "I bought b blue pencils which therefore cost 5xb pence altogether", etc.

Fig 5.6 Children's responses to Question 22.

Blue pencils cost 5 pence each and red pencils cost 6 pence each.
I buy some blue and some red pencils and altogether it costs me 90 pence.

If b is the number of blue pencils bought, and
if r is the number of red pencils bought,
what can you write down about b and r?

22(Level 4)	%
5b+6r = 90	10
Two correct pairs, of (6,10), (12,5), (18,0), (0,15).	1
b + r = 90	17
6b+10r=90 or 12b+5r=90	6

It was not possible to classify all the different responses that children gave, but of those that were coded b+r=90 and answers like 6b+10r=90 are probably the most interesting. Most of the children who wrote b+r=90 could cope with level 3 items, yet their answer only seems to mean "<u>b</u>lue pencils and <u>r</u>ed pencils cost 90 pence", which gives limited information and in which the letters are being used as objects. (b+r=90 could be read as "the number of blue and red pencils that I bought cost 90 pence", but this is still very much tied to the concrete reality of the question, and not a pure statement about numbers: the number b+r is not equal to 90.) Children who gave answers like 6b+10r=90 had found one correct pair of values for the numbers b and r (6 and 10), but instead of expressing this in a form that showed that b and r are numbers (b=6, r=10) switched to using the letters as objects: "6 blue pencils and 10 red pencils cost 90 pence".

Letter as Specific Unknown

The preceding three categories all describe ways of avoiding generalised arithmetic, by not using the letters as unknown numbers. The opposite is true of the present category, even though the idea of a specific unknown number is still rather primitive.

The use of a letter as a specific unknown has already been mentioned for items 9iv (n sided figure), 14 (r=s+t, r+s+t=), 5iii (e+f=8, e+f+g=), 13viii (simplify 3a-b+a) and 13v (simplify (a-b)+b). This usage is also required to solve 4ii and 4iii, but is not necessary for 4i:

Fig 5.7 Children's responses to Question 4.

4i (Level 2) Add 4 onto n+5	%	4ii(Level 3) Add 4 onto 3n	%	4iii(Level 4) Multiply n+5 by 4	%
n+9	68	3n+4	36	4n+20 or 4(n+5)	17
				4n+5 or 4xn+5	19
		7n	31	n+20	31
9	20	7	16	20	15

It may seem surprising that 4ii, in particular, turned out to be quite so difficult (level 3). The required answer, 3n+4, appears to be very simple but is also rather unsatisfactory:

nothing has really been done with the 3n and the 4, but children have to recognise that this is all that can be done to combine the elements, since n is an unknown. Many children seemed unwilling to accept this and instead gave the answer 7n or just 7, in which the elements that were meaningful (3 and 4) were "properly" combined but the letter was simply left as it was or ignored entirely.

The answers 7n and 7 belong to the category letter not used and the same applies in 4iii to the answers n+20 and 20. However, this approach is sufficient to answer 4i correctly, as long as the letter is retained in the answer.

4ii and 4iii both involve specific unknowns, but 4iii is significantly harder (level 4) because of its structural complexity: the operation +4 has to be applied deliberately and consciously to both elements of the expression n+5, but many children just "attached" the operation to the expression as a whole, which corresponds very closely to all that is required for the correct answer 3n+4 in 4ii, but which in this case produced ambiguous answers like 4xn+5 and n+5x4.

(It can be argued that such answers simply arise through a lack of familiarity with the appropriate notation -in this case brackets. However, this is difficult to sustain for children at the end of their 3rd year of secondary school -and for the 18% of 4th years who gave similar answers- particularly as the ambiguity can be resolved without using brackets. Also it should be said that children giving such answers performed well on the test, obtaining a mean score of 21 for 30 selected items compared to a mean for all children of 14. Thus a case can be made that ambiguous answers are given because the significance of brackets has not been understood, rather than brackets being unfamiliar. The findings of Kieran (1979) support this view.)

Letter as Generalised Number

Here the letter is seen as being able to take several values rather than representing a unique number. A distinction can be made between the idea of a letter taking several values in turn and a letter representing a set of values simultaneously, but the relevant items on the Algebra test are not sufficiently numerous or discriminating for this to be done. A (literally) far fuller investigation of the generalised number concept has been undertaken by Harper (1979), whose work is discussed in Chapter 14 , and who graphically describes generalised number as a "pregnant numeral". The notion of "simultaneous values" leads to the concept of a variable, but as far as the two Algebra items shown below are concerned it seems likely that it is sufficient to see the letters as taking several values in turn.

Fig 5.8 Children's responses to items 16 and 18ii.

16(Level 3)		18ii(Level 4)	
What can you say about c if c+d=10 and c is less than d?	%	L+M+N = L+P+N is Always Sometimes Never true. (when)	%
c<5 c=1,2,3,4 (systematic list) c=10-d	11 19 4	Sometimes, when M=P	25
Unsystematic list One value only (usually c=4)	1 39	Sometimes. Or M and P given a specific value Never	14 51

In Question 16 children most commonly found just one value for the letter c (though they may have been willing to find other values, if prompted). 18ii was adapted from Collis (1975b).

To solve the item, Collis argues that children have to realise that M and P can each take on many values and that some of these values may coincide. Collis goes on to argue that children who can conceive of the 'remote probability' of the values meeting are likely to be "at home with variables" (ibid, page 48); however, this use of the term variables seems to differ from how Collis uses it on other occasions (eg ibid, page 6, and 1974, page 6) and is much broader than the way it is being used in this chapter (see below).

Items 16 and 18ii were more difficult than many of the specific unknown items on the Algebra test, and it may the case that children can handle specific unknowns before they conceive of generalised numbers. However, it is perhaps more fruitful to regard these two ways of interpreting letters as complementary, as it seems likely that in the course of many algebra tasks children will flip from one interpretation to the other, depending on which is momentarily the more convenient. For example children might solve item 22 (blue and red pencils) by treating b and r as specific unknowns, but at the same time realise that the answer $5b+6r=90$ is true for several values.

Letter as Variable

It is difficult to find a precise meaning for the term variable. The common practice of using it as a blanket term for the letters in generalised arithmetic obscures the different meanings that children give to the letters, and is so broad as to make the term redundant.

The Concise Oxford Dictionary (1964) gives the mathematical meaning of variable as "indeterminate, able to assume different numerical values". This provides a useful starting point, although it does not seem to go beyond the ideas already present in seeing a letter as a generalised number.

Collis at one point invokes the idea of a relationship: the child who has reached the final stage of development of the ALC concept

> "now becomes capable of dealing with variables as such because he can hold back from drawing a final conclusion until he has considered various possibilities, an essential strategy for obtaining a relationship as distinct from obtaining a unique result" (Collis, 1974,p6).

Collis goes on to say that a child at this stage, given the formula V=LxBxH, would be able to discuss meaningfully the effect of various transformations on the formula, for example the effect on V if L were increased, B decreased and H held constant. It will be recalled (Chapter 2) that these ideas led to the item about an increase in the number of arcs and nodes on Draft 1 of the Algebra test and eventually to Question 19:

Fig 5.9 Question 19.

$a = b + 3$. What happens to a if b is increased by 2? ....

$f = 3g + 1$. What happens to f if g is increased by 2? ....

A key feature of both parts of Question 19 seemed to be that they involved second-order relations, or a relationship between relations. (In the case of 19i, this is of the most simple kind, namely that the relation between one pair of values of a and b, "a is 3 more than b", "is the same as" the relation between all other pairs of values, or put more succinctly "a is always 3 more than b".) The ability to work with second-order relations is seen by Piaget as an important aspect of formal operational thought (Inhelder and Piaget,1958; Malpas and Brown,1974) and it was decided to use this notion to define variables: in effect,

> letters are used as variables when a second-order relation is established between them.

The distinction between variables defined in this way, and letters used merely as specific unknowns or generalised numbers, can be illustrated by the different meanings that accrue to the relationship, 5b+6r=90, between the number of blue and red pencils in Question 22. With the letters regarded as specific unknowns, the relationship is simply a statement which is true for a particular pair of numbers. This statement is essentially static. When the letters are regarded as generalised numbers, 5b+6r=90 becomes a statement that is true for several, but still essentially isolated, pairs of numbers (6,10 12,5 0,15 18,0). This involves the idea that b and r can change, but does not of itself indicate how they change, for which it is necessary to compare the values in some way. A first step in such a comparison might be to order the values and establish a qualitative correspondence of the kind "as b increases r decreases" (illustrated by the left-hand diagram below). However, the analysis can be taken a step further (the right-hand diagram below), to establish a relationship like "the increase in b is greater than the (corresponding) decrease in r" (or "'12 is greater than 6' by more than '5 is less than 10'" etc). This is a second-order relation and describes the degree to which a change in one of the unknowns in 5b+6r=90 produces a change in the other.

Fig 5.10 Qualitative correspondence and second-order relation for Question 22.

Returning to Question 19, it turned out that neither part of the question correlated well enough with the rest of the test to be selected for one of the 4 levels. It may be that the question was too cryptic (compare "b is increased by 2" with "you add 2 more arcs"), with the result that children who might otherwise have coped with the notion of variables misinterpreted the question. The only answers that were accepted as correct for 19i and 19ii were answers of the type "a increases by 2", "f increases by 6" respectively, but it of interest to consider two other kinds of answers that the children gave, which were described on the marking scheme as "static-incorrect"(code 5) and "static-'correct'"(code 2). For both answer types the

Fig 5.11 Children's responses to Question 19.

	19i	19ii	19i	19ii
	$a = b+3$ $\Delta b = 2$ $\Delta a = ?$	$f = 3g+1$ $\Delta g = 2$ $\Delta f = ?$	%	%
Code 1	$\Delta a = 2$	$\Delta f = 6$	21	7
Code 2	STATIC — $a+2 = b+5$ $a+2 = b+3$ $a = b+1$ $a+2$.	"CORRECT" $f+g$ = ANYTHING $f = 3(g+2)+1$ $f = 3g+7$ $f = 3g-5$ $f+6$.	10	2
Code 5	STATIC — $a = b+5$ $a = 2b+3$ $a = 3b+3$	INCORRECT $f = 3g+3$ $f = 5g+\ldots$ etc	20	27

given relationship seemed to be interpreted in a static way (eg "this a is 3 more than this b", rather than "a is always 3 more than b") but the code 5 answers were far more primitive: here the "increase by 2" was effected by attaching 2 to the given relationship in a very direct and unpremeditated way:

a=b+3 became a=b+5 or a=2b+3 or a=3b+3, whilst f=3g+1 was changed to f=3g+3, f=5g+1, etc. In contrast, the code 2 answers seemed to express a genuine attempt to construct a new relationship out of the old (or, in the case of 19ii, at least showed the ability to cope with the structural complexity of 3g+1): the code 2 answer a+2=b+5 is maintaining the balance between the "old" a and b; in a+2=b+3 and a=b+1 a is "old" and b is "new"; the answer f=3g+7, though in some ways similar to the code 5 answer a=b+5, at least shows an awareness that f increases by 6 rather than 2.

Confirmation that Question 19 was being misinterpreted by children who were performing well on the test, was later obtained by examining the performance of all the children tested in the 1976 large-scale survey (1128, 961 and 731 2nd, 3rd and 4th years respectively, and also 103 5th years, making a total of 2923 children) on 30 items from the Algebra test that were selected to form the 4 levels. As can be seen from the table below, the overall performance of children giving code 2 answers to Question 19 was comparable to those answering the question correctly (code 1).

Fig 5.12 Mean score on 30 Algebra items for respondents to Question 19.

	19i			19ii		
Response Code	1	2	5	1	2	5
Response Frequency % (N=2923)	19	10	20	8	3	27
Mean No. of Items Correct (30 Items)	20.5	19.2	14.7	23.9	24.5	14.8

Note: the mean number of items answered correctly by children in the total sample was 14.1 .

Question 3 (Which is larger, 2n or n+2 ?) did correlate well with the other items on the Algebra test (though it proved to be very difficult: only 6% of 3rd years gave a correct conditional response, while 71% chose 2n and 16% wrote n+2 or "the same"). The relevance of a second-order relation to this question can best be explained by considering what happens to 2n and n+2 when specific values are chosen for n, for example n=4 and n=7. This gives the pairs 8,6 and 14,9 for 2n and n+2, from which the most obvious conclusion, which holds for each pair in turn, is that "2n is larger than n+2"(a first-order relation). However, by analysing the values in more detail, and in particular by considering what happens as n changes, it is possible to see a more complex relationship between 2n and n+2, of the sort "as n increases, the increase in 2n is greater than the increase in n+2" ($14-8 > 9-6$), or "as n increases, the difference between 2n and n+2 increases" ($14-9 > 8-6$). These are second-order relations, whose significance lies in the fact that they open up the possibility that for some smaller value of n the difference between 2n and n+2 may be decreased to zero ($n=2$) or even reversed ($n < 2$).

The argument being advanced here is not that children go through precisely these steps to solve Question 3, but rather that children who are able to cope with complex relations of this sort (through having sufficient "processing capacity" perhaps) are likely to consider the possible effect of n on on the relative size of 2n and n+2, whereas other children (with less processing capacity) will go for something simpler and more direct.

None of the other items on the test can be said to involve second-order relations, although the coordinations required to solve Question 21 are perhaps equally complex. Here it is necessary to realise that a set x can equally well be represented by $5x$, and furthermore that the resulting transformation on the values of x is $\div 5$, the inverse of x5. The Question was answered correctly by 12% of 3rd years.

Fig 5.13 Question 21.

If this equation ⟶ is true when $x = 6$, $\qquad (x + 1)^3 + x = 349$

then what value of x will make this equation ⟶ true? $\qquad (5x + 1)^3 + 5x = 349$

$x =$ ....

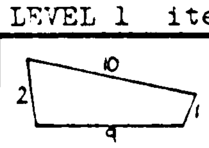
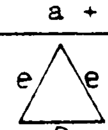
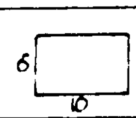
Levels of Understanding in Algebra

Items from the test that correlated well with each other were classified into 4 "levels of understanding". In the case of Question 13, only 5 of its 9 items were chosen, not because the others correlated badly but to avoid having too many items of a very specific kind. Altogether 30 of the 51 items on the test were selected. In what follows children are described as "being at" a given level if they correctly answered about two thirds of the items at that and no higher level. (Infact the criterion was 4/6, 5/7, 5/8 and 6/9 items correct, for Levels 1 to 4 respectively.)

Level 1

The items for this level are shown in abbreviated form below, together with their 3rd year facilities. As can be seen, the items were very easy.

Fig 5.14 Items assigned to level 1.

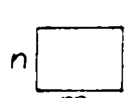
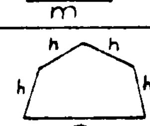
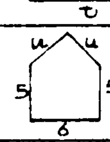
facility	and item	LEVEL 1 items	other responses	interpretation of letter deemed adequate for correct response
97	8	 p=		no letters involved
97	5i	If $a + b = 43$ $a + b + 2 = \ldots$		NOT USED
94	9i	 p=		OBJECT
92	6i	What can you say about a if $a + 5 = 8$?		EVALUATED
89	7ii	 A=		no letters involved
86	13i	$2a + 5a =$		OBJECT

The items at level 1 are purely numerical (8 and 7ii), or they have a simple structure and can be solved by using the letters as objects (9i and 13i), or by evaluating the letter (6i), or by not using the letters at all (5i). For more complex items, children at this level tended to give answers like 4ht or 5ht instead of 4h+t for item 9ii, 8ab instead of 3a+5b for 13iv, 763 instead of 761 for 5ii. Where items required specific unknowns, these children were likely to evaluate the letter (p=32 instead of p=2n for 9iv, e+f+g=12 or 9 instead of 8+g for 5iii), or they did not use the letter at all (7 or 7n instead of 3n+4 for 4ii).

Level 2

The clear difference between these items and those of Level 1 is their increased complexity, though the letters still only have to be evaluated (11i and 11ii below) or used as objects (9ii, 7iii, 9iii and 13iv).

Fig 5.15 Items assigned to level 2.

		LEVEL 2 items	other responses		interpretation of letter deemed adequate for correct response
75	15i	A shape with 57 sides has ... diagonals (given rule "take 3")			no letters involved
68	7iii	 A=			OBJECT
68	9ii	 p=	4ht or hhhht	20	OBJECT, but need to avoid temptation to close (4ht)
64	9iii	 p=	2u16 or uu556	16	OBJECT, but need to avoid temptation to close (2u16)
62	11ii	What can you say about m if m = 3n + 1 and n = 4			EVALUATED, but need to cope with (temporary) ambiguity of m = 3n + 1
61	11i	What can you say about u if u = v + 3 and v = 1	u=2	14	EVALUATED, but need to cope with ambiguity of u=v+3 (one unknown is 3 more than another unknown), and not reduce this ambiguity to "u and v together equal 3" which leads to u=2.
60	13iv	2a + 5b + a =	8ab	20	OBJECT, but need to avoid temptation to close (8ab, which is also ambiguous)

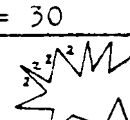
It might be argued that the advance made at this level (eg being able to write 2u+16 instead of 2u16, 3a+5b instead of 8ab) is due simply to an increased familiarity with algebraic notation. However, when the Calvert raw scores of a subsample of 2nd year children was examined, it was found that children using the correct notation had substantially higher mean scores, which suggests that the use of correct notation is at least in part conceptual. (For the answers 2u+16 and 2u16 the mean Calvert raw score was 51 and 46 respectively, and for 3a+5b and 8ab the means were 53 and 44. The mean for all children in the subsample, N=610, was 46.)

The advance made at level 2 can also be seen as a greater willingness(much more fully realised at level 3) to accept answers that appear incomplete or ambiguous (Collis's ALC). For example, it may be the case that many of the children who gave the answer 8ab as a simplification of 2a+5b+a (of whom about three quarters were at level 1) knew how to write 3a+5b but preferred their answer because it looked more complete. In 11i most level 1 children omitted the item or gave the answer 2 instead of 4 (u=v+3, v=1, u=). Here it may be the ambiguity of u=v+3 (one unknown is 3 more than another..) that caused these responses, either putting children off entirely or prompting them to read the equation as "u plus v equals 3", which reduces the ambiguity but leads to the answer 2.

Level 3

This level signals a major step forward, in that for the first time letters are genuinely being used as unknown numbers (specific unknown). Answers like 8+g and 3n+4 are now meaningful, despite their lack of closure and even though the letters represent numbers rather than objects. However, children at this level can only cope with specific unknowns when the structure of the items is simple: for example, they can cope with "Add 4 onto 3n", where the operation +4 is simply attached to what is given, but with "Multiply n+5 by 4" they have difficulty coordinating the operations and are likely to give an ambiguous answer like n+5x4 .

Fig 5.16 Items assigned to level 3.

		LEVEL 3 items	other responses		interpretation of letter deemed adequate for correct response
52	15ii	A shape with k sides has ... diagonals			SPECIFIC UNKNOWN
47	13viii	3a - b + a =			OBJECT/SPECIFIC UNKNOWN
45	13ii	2a + 5b =	7ab or 12a	34	OBJECT, but strong temptation to close.
41	5iii	If e + f = 8 e + f + g = ...	12 9 15	26 6 2	SPECIFIC UNKNOWN. Though e and f can be ignored by matching (as in 5i) g can not; nor can g be evaluated (4+4+4=12, etc).
41	14	What can you say about r if r = s + t and r + s + t = 30	r=10	21	SPECIFIC UNKNOWN. r can not be evaluated directly (as in 10+10+10=30).
38	9iv	Part of this figure is not drawn...  n sides altogether. p=	p=32, 34, etc	18	SPECIFIC UNKNOWN. n can not be used as an object (as in 9i, 9ii, 9iii), nor can n be evaluated (by literally closing the figure: p=32 etc)
36	4ii	4 added to n can be written as n+4. Add 4 onto 3n.	7n 7	31 16	SPECIFIC UNKNOWN. n has to be operated upon, not avoided(4+3n → 7n) or ignored entirely(4+3n → 7).
30	16	What can you say about c if c + d = 10 and c is less than d	only one value, usually c=4	39	GENERALISED NUMBER

The first (and easiest) three items shown in the diagram should perhaps be regarded as being at the onset of level 3, rather than fully within it. In the case of 15ii, though k is a specific unknown, the operation -3 is given explicitly and there is really little that children can do other than write k-3: for example, there are no other numbers acting as distractors, as in 4ii where many children succumb to combining the 3 and the 4. (Despite this, however, 17% of 3rd year children avoided using a specific unknown by evaluating k, many of them using the alphabet code to get answers like h or 8 by counting back from k or 11.) Item 13viii (simplify 3a-b+a) might be said "just" to require specific unknowns, in that the interpretation of the letters as objects, which is perfectly adequate for an expression like 2a+5b+a (item 13iv), is no longer fully plausible because of -b. (Put another way, the subtraction increases the level of ALC.) In 13ii the letters can be regarded as objects, but children have to overcome the temptation to do something to the expression 2a+5b (simplify if possible..). Not surprisingly many children (34%) gave answers like 7ab or 12a (where one b is exchanged for two a's).

Question 16 is the only item in the group that requires the use of letters as generalised numbers. The only reason it is classified at level 3 rather than level 4 is because its facility pattern, over the three year groups tested, is closer to the level 3 items.

Level 4

At this level children can cope with specific unknowns in items having a complex structure (13v, 4iii, 7iv). They can also cope with items like 20, 22 and 17i which require, at a minimum, that the letters are regarded as specific unknowns but where there is a strong temptation to treat the letters as objects (cakes and buns, pencils, wages and hours). 18ii involves generalised numbers, whilst item 3 involves variables.

Fig 5.17 Items assigned to level 4.

		LEVEL 4 items	other responses	interpretation of letter deemed adequate for correct response
25	18ii	Is the following always, never or sometimes(when) true? $L + M + N = L + P + N$	Never 51	GENERALISED NUMBER. M and P can represent a range of values, which may coincide.
23	13v	$(a - b) + b =$		SPECIFIC UNKNOWN. The use of letters as objects is no longer plausible (an apple take away a banana..).
22	20	Cakes cost c pence each and buns b pence each. If I buy 4 cakes and 3 buns, what does $4c + 3b$ stand for?	4 cakes and 3 buns 39	SPECIFIC UNKNOWN (or generalised number). The temptation to use the letters as objects is particularly strong, since the item involves objects.
17	4iii	Multiply $n+5$ by 4.	4xn+5 17 n+20 31 20 15	SPECIFIC UNKNOWN. Here it is necessary to coordinate two operations, and to recognise the ambiguity of an answer like 4xn+5 .
12	7iv	5 ▭ e 2 A=	5xe+2 18 e+10 28 10 13	SPECIFIC UNKNOWN.
12	21	If $(x+1)^3+x = 349$ is true when $x=6$, what value of x makes $(5x+1)^3+5x = 349$ true?		GENERALISED NUMBER or variable. x can be represented by 5x, which results in the transformation ÷5.
11	22	b blue pencils (5 pence each), r red pencils (6 pence each), cost 90 pence altogether... (ie $5b + 6r = 90$)	b+r=90 17 6b+10r=90 6 or 12b+5r=90	SPECIFIC UNKNOWN or generalised number. Not letter as object (blue and red pencils cost 90 pence, etc).
6	3	Which is larger, 2n or n+2 ? Explain.	2n (because it's multiply, etc) 71	VARIABLE(2nd order relationship). Intuitively it is reasonable to assume $2n > n+2$ (eg for n=10, 20>12). But as n changes the difference between 2n and n+2 changes, so for some (smaller) value of n 2n may be less than n+2
5	17i	Basic wage £20. £2 per hour of overtime. W, total wage; h, number of hours of overtime. Equation in W,h ($W=20+2h$).	W=20+h 13 W+h 14 20w+2h 11	SPECIFIC UNKNOWN or generalised number. Not letter as object (W=20+hours overtime).

Summary: the Items at each Level

The items at levels 1 and 2 can be solved without having to operate on letters as unknowns; instead the letters can be evaluated, not used, or regarded as objects. At levels 3 and 4 the letters have to be treated as specific unknowns, generalised numbers or variables.

The difference between level 1 and level 2, and between level 3 and level 4 is essentially a matter of complexity. For example, in the level 1 item "Find a if a+5=8", the letter can be evaluated immediately, by recalling a familiar number bond, whereas in the level 2 item "Find u if u=v+3 and v=1" it is first necessary to cope with an ambiguous statement. And whereas in the perimeter item 9i the objects being collected together are all of the same type (p=3e), in 9ii (level 2) the objects differ, which means that the answer (p=4h+t) cannot be closed. Similarly whilst the level 3 item "Add 4 onto 3n" essentially involves just a single operation, in the level 4 item "Multiply n+5 by 4" the operations +5 and x4 have to be coordinated.

Summary: the Children at each Level

The table below shows how many children at each level gave answers classified by the codes 0 to 9 for item 4ii (for the total 1976 sample, N=2923).

Fig 5.18 Level of respondents to item 4ii.

Responses to Item 4ii	Level of Children: Level 0	Level 1	Level 2	Level 3	Level 4	Row Total
Code 1	5	93	130	562	181	971
Code 3	5	35	39	26	4	109
Code 4	0	3	0	2	0	5
Code 5	52	509	309	125	1	996
Code 6	55	250	117	17	2	441
Code 8	20	73	31	3	0	127
Code 9	16	94	26	12	1	149
Code 0	49	61	10	5	0	125
Column Total	202	1118	662	752	189	2923

The raw frequencies are not easy to interpret: for example, whilst it is true to say that most children at level 4 gave code 1 answers, it is not the case that most children who gave code 1 answers were at level 4. Also the differences in performance at different levels and for different codes are not sufficiently clear cut to be able to make statements like "all children at level 3 and above (the level of the item)

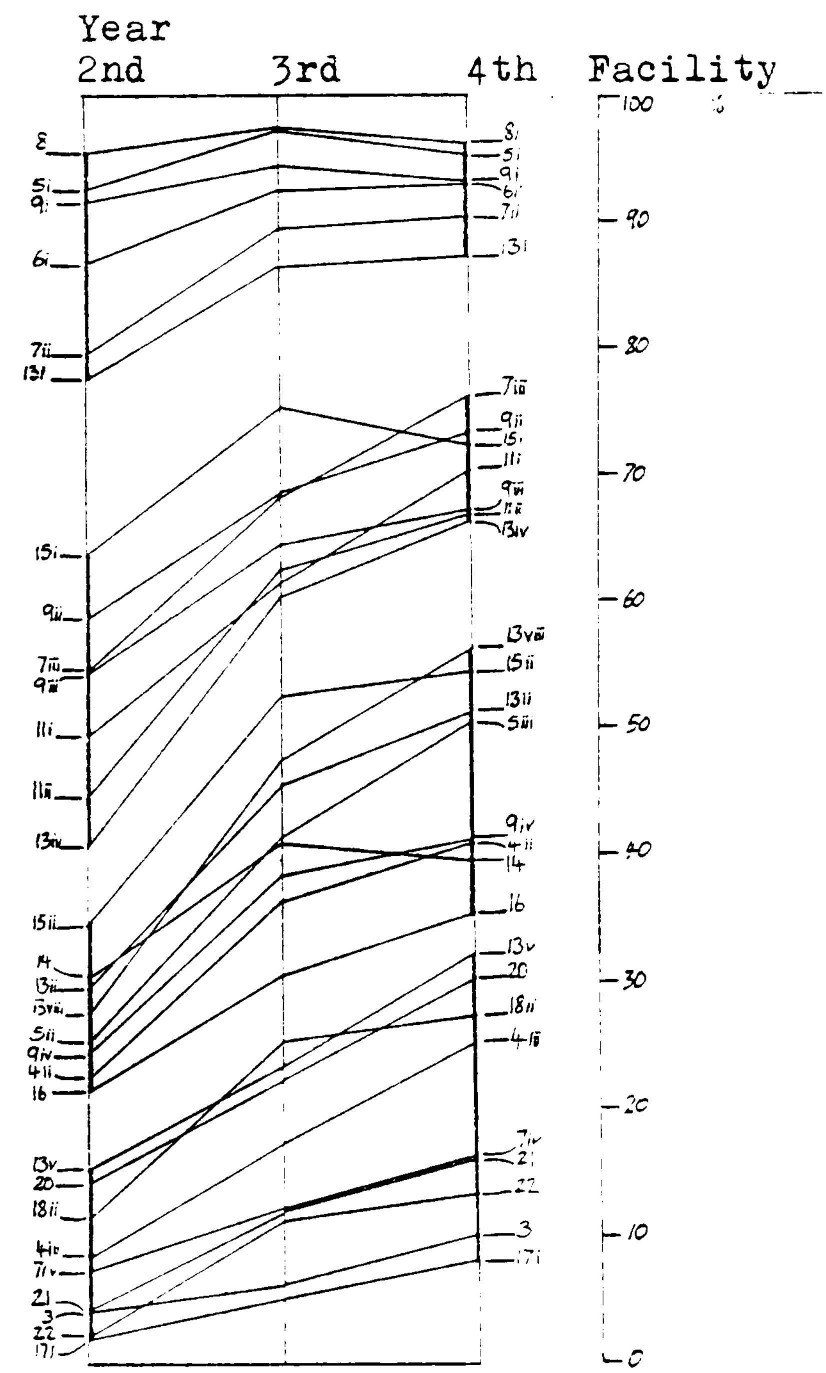
Chapter 6 CHANGES IN PERFORMANCE WITH AGE

Cross-Sectional Data: Facility of Items

The facilities of the 30 items that were selected to form the levels are shown in the table and graph below, for the 1976 2nd, 3rd and 4th year samples (13, 14 and 15 year olds).

Fig 6.1 Facility of 30 selected Algebra items for 1976 2nd, 3rd and 4th year samples.

	Item	Year 2	3	4th
Level 1	8	95	97	96%
	5i	92	97	95
	9i	91	94	93
	6i	86	92	93
	7ii	79	89	90
	13i	77	86	87
Level 2	15i	63	75	72
	9ii	58	68	73
	7iii	54	68	76
	9iii	54	64	67
	11ii	44	62	67
	11i	49	61	70
	13iv	40	60	66
Level 3	15ii	34	52	54
	13viii	27	47	56
	13ii	29	45	51
	5iii	25	41	50
	14	30	41	39
	9iv	24	38	41
	4ii	22	36	41
	16	21	30	35
Level 4	18ii	11	25	27
	13v	15	23	32
	20	14	22	30
	4iii	8	17	25
	7iv	7	12	16
	21	4	12	16
	22	2	11	13
	3	4	6	10
	17i	2	5	8



A brief examination of the graph is sufficient to show that the increase in item facility was generally greater from the 2nd to 3rd year sample than from the 3rd to 4th years. In some cases performance actually declined from the 3rd to 4th years, answered the item correctly'. However, by taking into account the size of the marginals (ie by considering row and column percentages, as in the table below) it is possible to make statements such as the following:

Most children (51%) who gave code 5 answers were at level 1.
Most children (57%) who gave code 6 answers were at level 1.
Most children (58%) who gave code 8 answers were at level 1.

Children at level 1 most commonly (46%) gave code 5 answers.
Children at level 1 frequently (22%) gave code 6 answers.
Children at level 1 rarely (7%) gave code 8 answers.

Children at level 2 most commonly (47%) gave code 5 answers.

Fig 5.19 Percent of code 5, 6 and 7 respondents to item 4ii at each level and percent of level 1 and 2 children giving different responses to 4ii.

Responses to Item 4ii	Level 0	Level 1	Level 2	Level 3	Level 4	
Code 1		8	20			
Code 3		3	6			
Code 4		0	0			
Code 5	5	51 46	31 47	13	0	100
Code 6	13	57 22	27 18	4	1	100
Code 8	16	58 7	24 5	2	0	100
Code 9		8	4			
Code 0		6	2			
		100	100			

On the basis of these statements (and similar ones derived from the cross-tabulations shown in Appendix 5), the table below has been constructed, which lists some of the more "typical" responses of children at each level.

Fig 5.20 Summary of performance at each level.

Items									
Level	Level 1		Level 2			Level 3		Level 4	
Structure	simple		complex			simple		complex	
Required Use of Letters	evaluated not used object					specific unknown generalised number		variable	
	9i e e e	6i a+5=8	11i u=v+3 v=1	9ii h h h h t	13iv 2a+5b +a	5iii e+f =8 e+f+g=	4ii Add 4 to 3n	4iii X n+5 by 4	22 pencils
Level 1 Response	✓ OBJCT	✓ EVAL	omit. u=2 EVAL	4ht OBJCT	8ab OBJCT	12 9 EVAL	7 7n NOT U	20 n+20 NOT U	-
Level 2 Response			✓ EVAL	✓ OBJCT	✓ OBJCT	12 EVAL	7n NOT U	n+20 NOT U	-
Level 3 Response						✓ S UNK	✓ S UNK	4xn+5 S UNK	b+r=90 OBJCT
Level 4 Response								✓ S UNK	S UNK GEN or VAR

particularly for some of the easiest (level 1) items. In part this is probably due to a ceiling effect, but it is also likely that this reflects a growth in the proportion of children who reject mathematics, especially amoungst those who find the subject difficult. For the harder items the 4th year children's performance shows a steady improvement over that of the 3rd years, as can be seen from the first table below. However, both tables again show very clearly that the improvement between the 2nd and 3rd year sample was far greater.

Fig 6.2 Average increase in facility from 2nd to 3rd to 4th years for items at each level, and number of items correct.

Average increase in facility

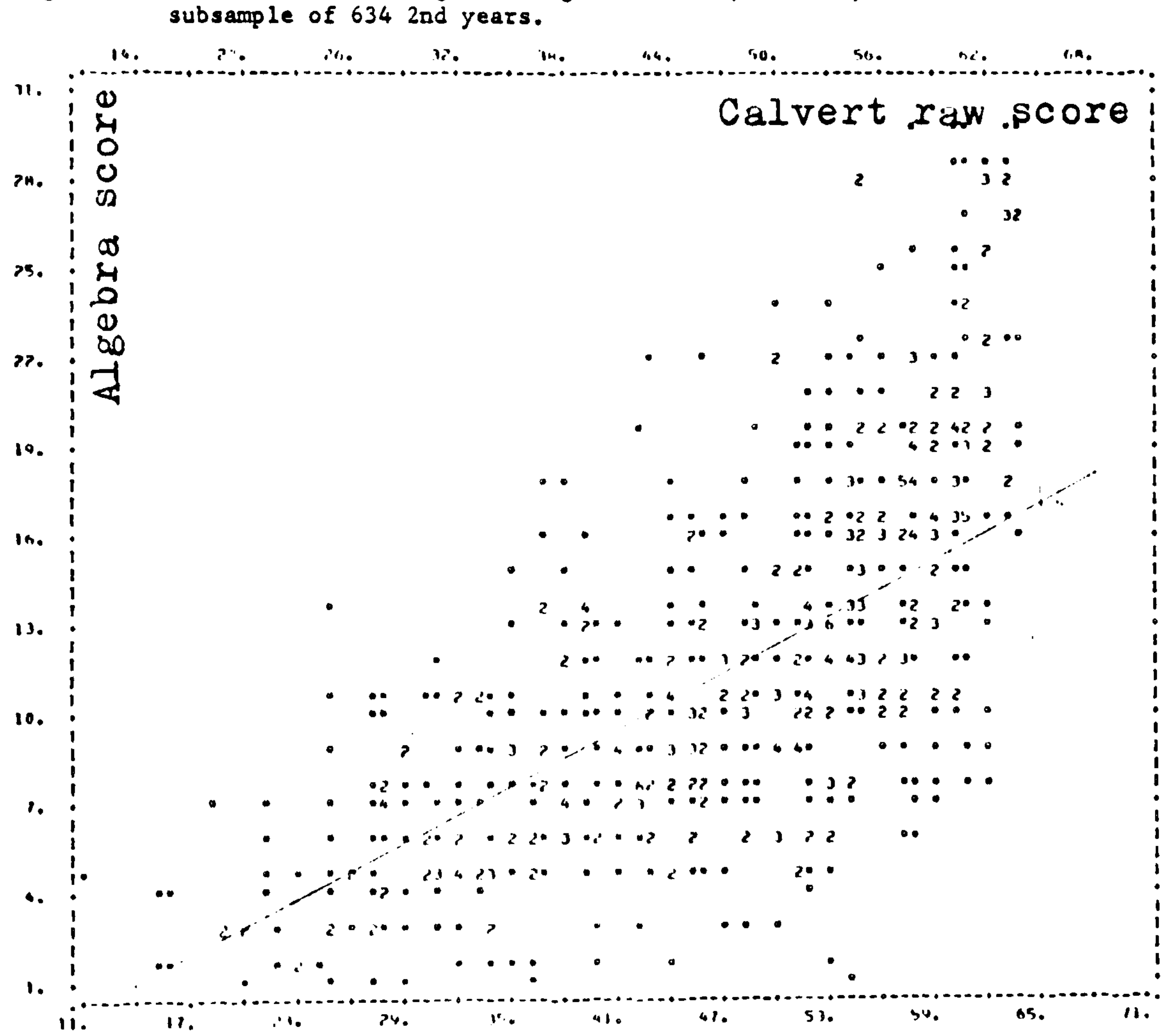
Item level	1	2	3	4
2nd year to 3rd year	6	14	15	7%
3rd year to 4th year	0	5	5	5%

Number of items correct (30 items)

	Mean	Increase
2nd year	11.6	
		3.15
3rd year	14.7	
		1.13
4th year	15.9	

A possible explanation for the relatively poor performance of the 2nd year children is that they may have been handicapped by a lack of familiarity with generalised arithmetic; on the other hand, once a minimum level of familiarity has been reached (in the 3rd year) it may be the case that performance is far more dependent on cognitive ability, than on the experiences of algebra that children typically meet at school. The product-moment correlation between total score on the 30 selected Algebra items and Calvert raw score was found to be 0.69 for a sub-sample of 634 2nd years (despite a marked "topping effect" on the Calvert test, which is demonstrated by the scattergram below). The value of the correlation suggests that cognitive ability does have a substantial effect on performance, even for 2nd year children. At the same time, the view that they suffered from a lack of familiarity is supported by informal observations made at the time the class test was being developed: many 2nd year children found the test a puzzling and uncomfortable experience, in contrast to the confidence generally shown by the older children even when their performance turned out to be low. Also, an examination of 1st and 2nd year text books (for example the SMP lettered series) suggests that children's early experiences of algebra are presented with little purpose or conviction (see Chapter 15).

Fig 6.3 Calvert raw score against Algebra score (30 items) for a subsample of 634 2nd years.



Cross-Sectional Data: Distribution of Levels

It was mentioned briefly in Chapter 5 that children were assigned to one of the four Algebra levels according to the highest level at which they answered at least two-thirds of the items correctly (strictly, 4/6, 5/7, 5/8 and 6/9 for the level 1, 2, 3 and 4 items respectively). The tables below show the distribution of 2nd, 3rd and 4th year children at each level, for the 1976 samples. (These consist of 2820 children, but 34 children have been excluded from the tables because their performance did not scale, ie they reached criterion at one level but not at all lower levels.)

Fig 6.4 Percent and cummulative percent of children at each level.

Percentage of children at each level

Child level	0	1	2	3	4
2nd year	10	50	23	15	2
3rd year	6	35	24	29	6
4th year	5	30	23	31	9

Cummulative percentage of children at each level

Child level	0+	1+	2+	3+	4
2nd year	99	89	40	17	2
3rd year	99	93	58	34	6
4th year	99	93	63	40	9

As is to be expected, this information presents the same overall picture as the facility comparison, with the 2nd years performing substantially worse than the 3rd years, and the 4th years only slightly better. However, as far as the improvement of the 4th years at least is concerned, assessing performance in terms of levels seems to obscure the gain made in the number of items answered correcly, since the difference between the 3rd and 4th year means (1.13) is highly significant ($t=3.33$, $p \simeq 0.001$), but the difference between the distribution of levels is not ($\chi^2=5.93$) despite the large numbers of children involved.

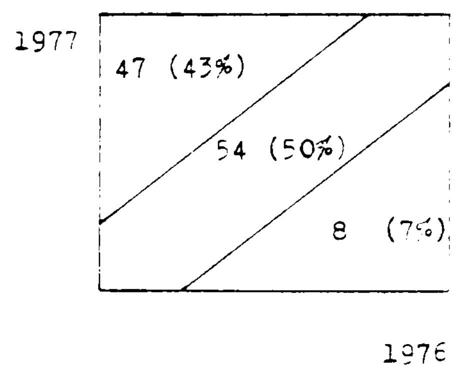
Longitudinal Data

A group of 200 2nd year children (from schools 05, 06, 07 and 16), who formed part of the 1976 survey, were tested again in the summer of 1977 and 1978. Equal numbers of children were selected randomly from four IQ groups (standardised Calvert score) so as to give a fairly rectangular, rather than normal, distribution. Due to children being absent or moving school, only 109 of the original 200 children were tested on each of the three occasions, with the drop out rate being slightly higher amoungst the children with low IQ.

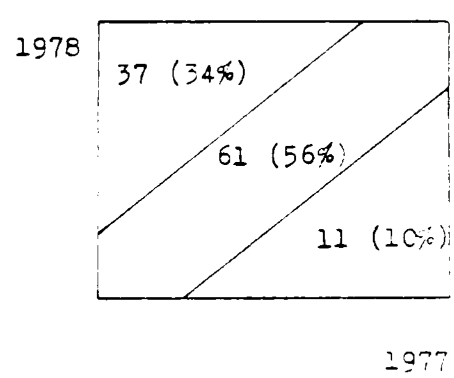
Children's performance was classified into levels P1, P2, P3 and P4 which are comparable to Levels 1, 2, 3 and 4, but which were determined from a slightly different subset of items (PIG 8). The change in the number of children at each level from one year to the next is shown in the tables below.

Fig 6.5 Levels (PIG 8) of 109 children in 1976, 1977 and 1978.

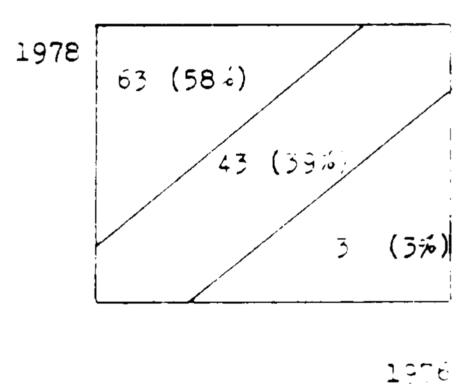
	5	34	57	11	2	109
1977 P4			1	1	1	3
P3			19	7		26
P2	3	21	34	3	1	62
P1	2	12	3			17
P0		1				1
	P0	P1	P2	P3	P4	
					1976	



	1	17	62	26	3	109
1978 P4			4	8	2	14
P3			14	18	1	33
P2		10	36			46
P1	1	5	7			13
P0		2	1			3
	P0	P1	P2	P3	P4	
					1977	



	5	34	57	11	2	109
1978 P4			6	7	1	14
P3		4	24	4	1	33
P2	1	18	27			46
P1	3	10				13
P0	1	2				3
	P0	P1	P2	P3	P4	
					1976	



The most important aspect of these data is that the proportion of children whose performance regressed is small, with a maximum of 10% over one year and only 3% over two years. This suggests that it would be reasonable to interpret the cross-sectional data from the 1976 survey as a picture of how children's performance progresses as they get older.

The first two graphs below show the change in level over two years for each child in the longitudinal study, against their Calvert standardised and raw score respectively. The graphs show that there is a strong relation between the levels and Calvert score. The third graph, which is a simplification of the second, suggests there may also be a relation between change in level and Calvert score, with a greater proportion of the children at the top end of the Calvert range changing levels over the two year period (and perhaps also a greater proportion changing by more than one level). However, this has to be interpreted with caution: although it is reasonable that children at the higher levels (and with the higher Calvert scores) should progress through the levels more quickly (otherwise, how do they get there?), it happens to be the case that at all three phases of the survey over half the children are at level P2, so that the greater degree of change amoungst those with high Calvert scores might be the result of their having been at P2 for some time, whilst those at the middle of the Calvert range might just be entering P2 or consolidating their position within it. In addition, it is possible that children take longer to traverse P2 than the higher levels. Unfortunately, a more extended longitudinal survey is probably required to answer these questions satisfactorily.

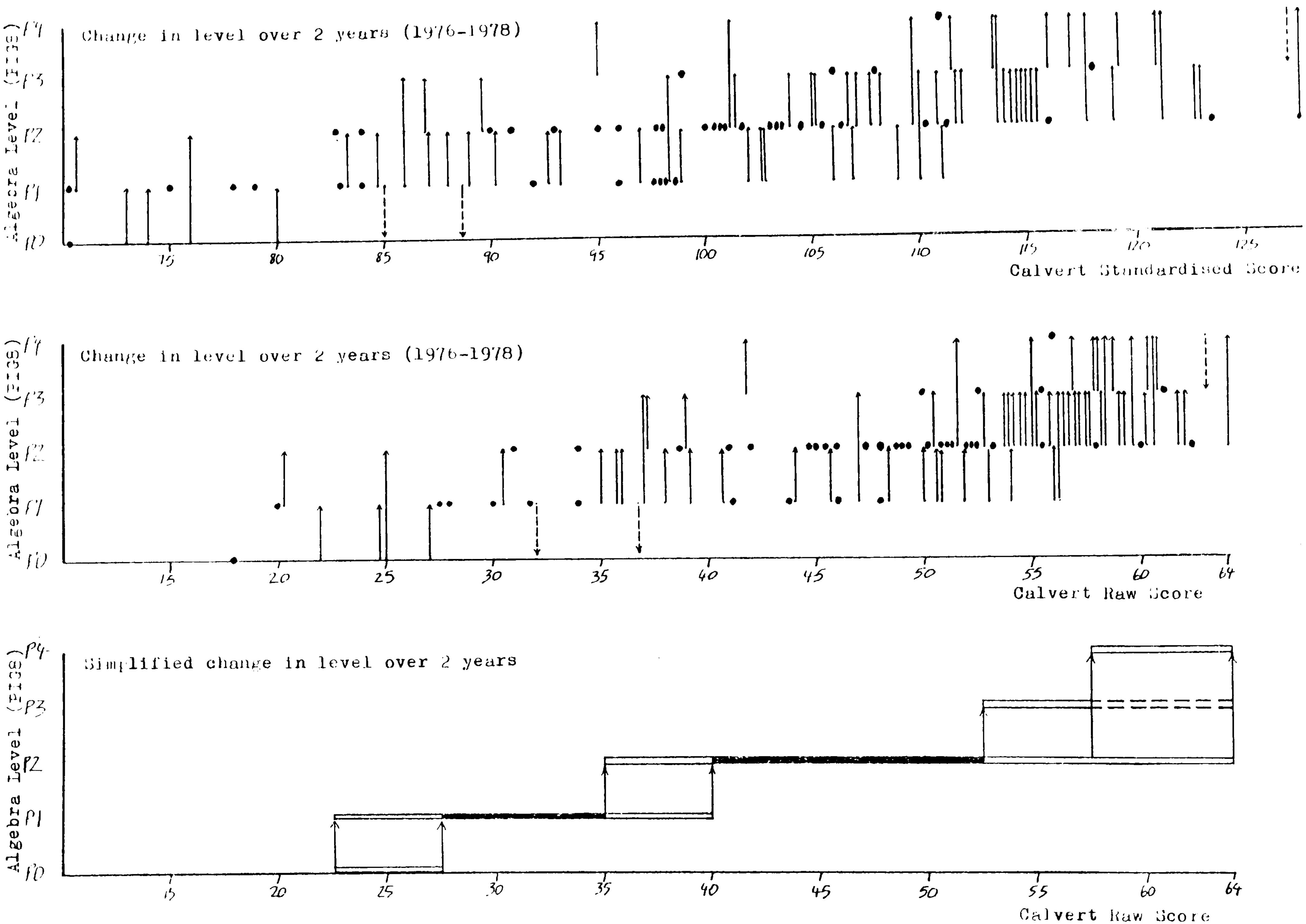


Fig 6.6 Change in levels over 2 years against children's Calvert standardised and raw scores, and simplified change x raw scores.

Chapter 7 THE STATISTICAL ANALYSIS OF THE ALGEBRA TEST: OVERVIEW

The classification of items into levels (and subsequently assigning children to these levels) was undertaken for each of the CSMS mathematics tests (and also for the science tests, where the levels were explicitly described in Piagetian terms). This procedure can be regarded simply as a device for organising the research findings into a more readily communicable form. However, as was stated in Chapter 1, in adopting the procedure for the Algebra test it was also assumed, at least as a first approximation, that children all go through the same stages in their grasp of a given concept and that their level of understanding is consistent across different concepts.

With respect to the items of the Algebra test, these assumptions implied that children who could cope with one item should be able to cope with all other items of the same and higher facility, and that similarly, once items had been classified into levels, children who could cope consistently with the items at one level should be able to cope at all the easier levels. The aim of the statistical analysis was to identify the items that satisfied these criteria most closely. However, first a method of analysis had to be found for doing this, which proved to be an extremely difficult and time consuming task. Initially a method based on item characteristic curves (eg Guilford,1965,p483) was tried, that had been developed by the science wing of CSMS (Shayer,1978a,p31-37; Adey,1980,p27) and which is discussed in the next chapter. The method had a number of virtues, not least that it allowed close contact to be maintained with the data, but it was not entirely satisfactory. It gradually emerged that three related problems needed to be solved: first a suitable measure of association was required, in order to assess the similarity or hierarchical nature of pairs of items; then a method of analysis needed to be found so that this measure could be used to assess the homogeneity of several items or of the test as a whole; finally, having selected the most homogeneous items and grouped them into levels, it was felt that a method was needed to check that the levels formed a scale, ie to assess the extent to which children who passed one level also passed the easier levels.

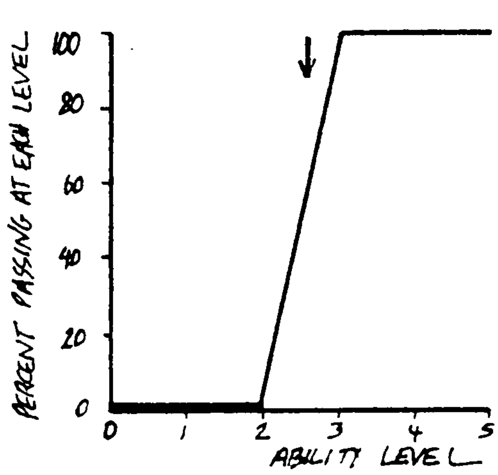
The measure of association that was finally chosen was the phi coefficient, which is discussed in Chapter 9. Chapter 10 discusses the way in which phi was used to analyse the test as a whole: in the event two methods were used, one of which was based on rather intuitive drawings that became known as "spider diagrams", whilst the other involved factor analysis. Guttman scalogram analysis was used to assess the consistency of the levels: initially the method seemed ideally suited to this purpose and was greeted with enthusiasm, but a number of shortcomings have since come to light which are discussed in in Chapter 11.

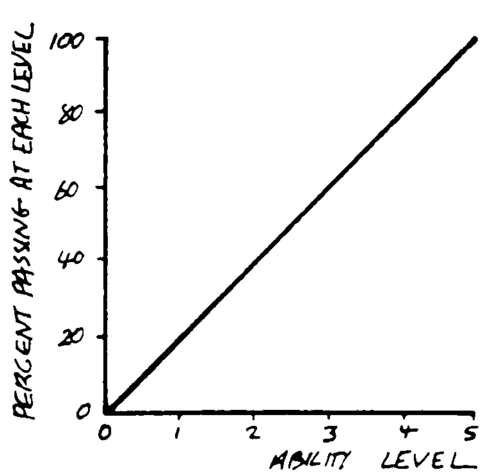
If desired, these chapters on the statistical analysis of the test (PART B) can be read after the final section of the thesis (PART C).

Chapter 8 CHARACTERISTIC CURVES

For a test that measures along a single scale of ability, a characteristic curve provides a way of displaying the degree of discrimination of an item and the point (or level) on the scale at which the item discriminates best. Two such curves are shown below, for imaginary items of average facility from a test where ability is measured on a 6 point scale. The first item has perfect discrimination, with all subjects from level 3 onwards passing the item and all those with lower ability failing the item; in contrast the second item discriminates along the whole ability scale, with higher ability subjects having a greater probability of passing the item, but its discriminating power is far less. According to Guilford (1965,p484) a test consisting of items whose curves come close to that of the first item is more effective at grading individuals, but only if the items span a wide range of facility; unfortunately, however, "With the extensive range of difficulty level, there could not be as high internal reliability as some might desire".

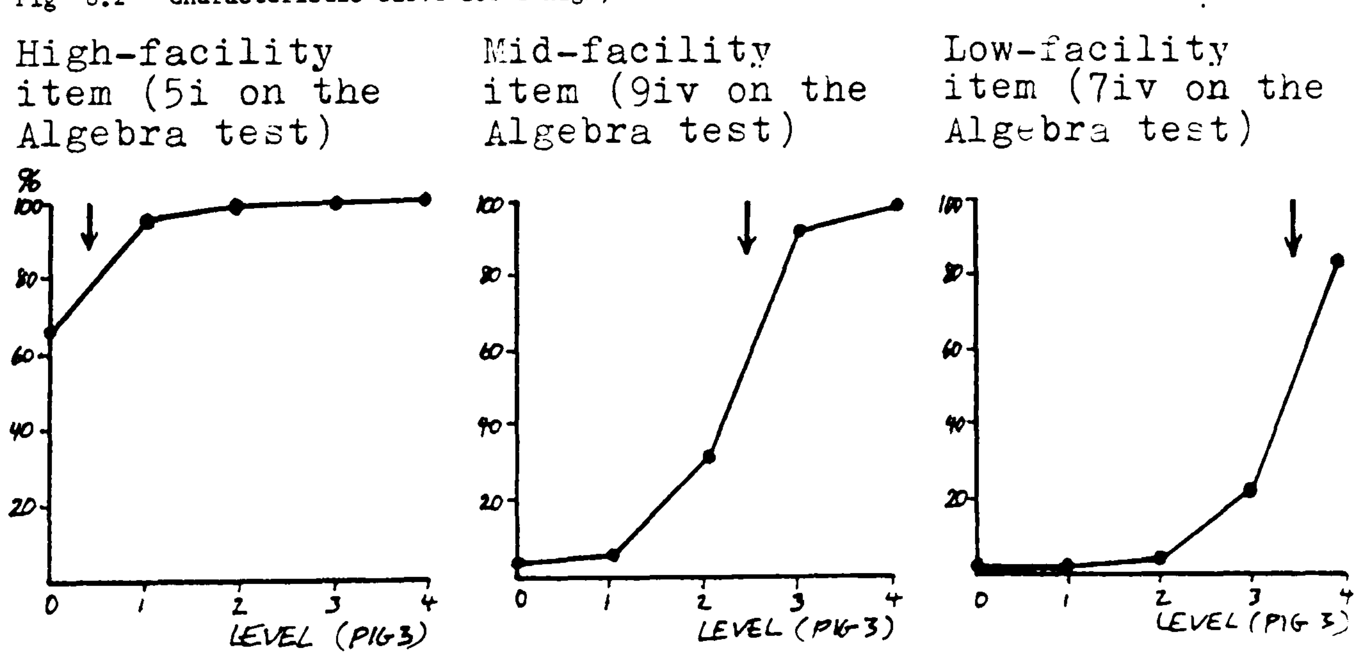
Fig 8.1 Ideal and more usual item characteristic curves.





In practice, tests are usually constructed with the items all having a facility of about 50% and with a discrimination pattern similar to that of the second example above (if such equal-facilit items all discriminated perfectly, the resulting test would partition subjects into just two groups, and would be no more effective than any one of the items on its own). In contrast, the CSMS tests (mathematics and science) were deliberately designed to have items with widely differing facilities, not only to discriminate sharply between children at different levels of ability, but so that children's levels of ability could be described qualitatively, in terms of the items actually passed instead of just in terms of a numerical score. Thus the characteristic curve of the first item shown above provided an appropriate model against which to compare the CSMS items. The diagram below shows such curves for 3 items from the Algebra test, of high, medium and low facility respectively (the levels are from one of the early scales into which the test was analysed, PIG 3). These items were regarded as discriminating well, though compared to the ideal the curves have a far more rounded, S-shaped appearance, and in the case of the extreme-facility items only a partial S-shape.

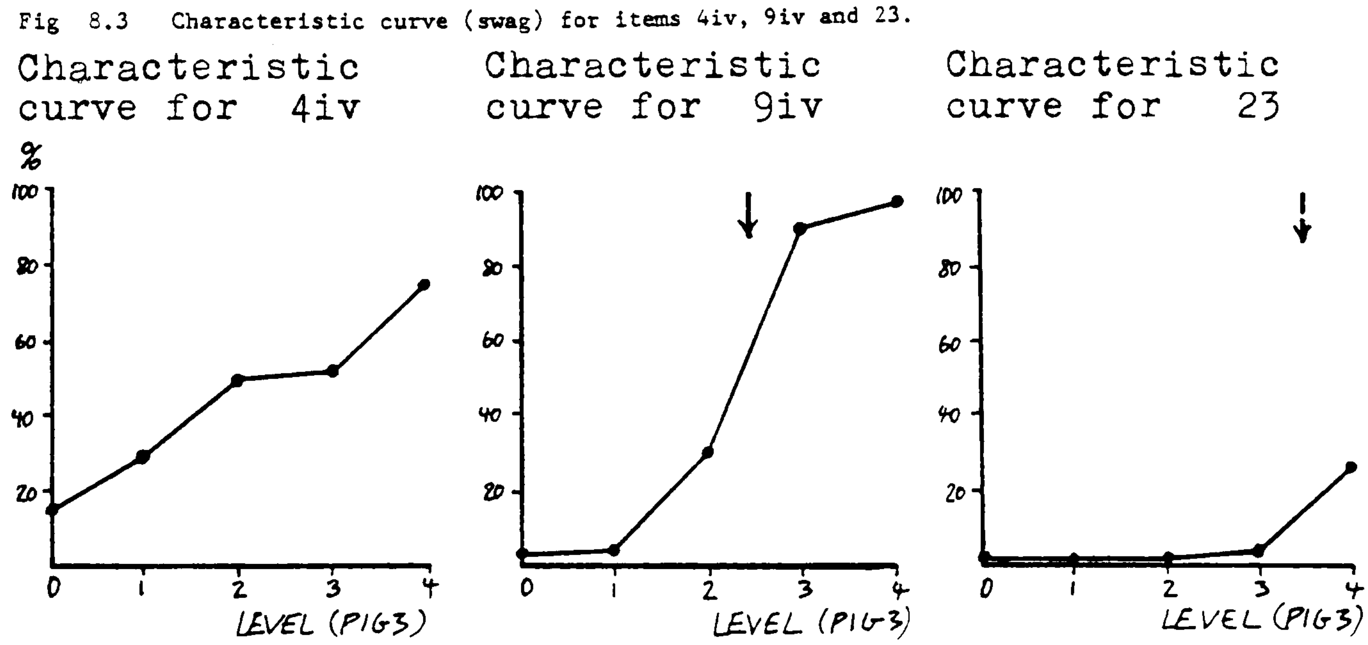
Fig 8.2 Characteristic curve for a high, medium and low facility item.



The science wing of CSMS used characteristic curves in the method they developed for analysing their Piagetian class tasks. The method is an iterative one, in which the first step is to classify the items on a test into levels (in the case of the science wing this classification was based on Piaget's descriptions of the tasks that their items were designed to model). Next, scoring rules are devised, and children assigned to levels according to their performance on the tests (the most common rule, at least initially, was to assign children to the highest level for which at least two-thirds of the items had been answered successfully). Characteristic curves, or "swags" as they affectionately became know (after Shayer, Wylam and Guilford), are then drawn for each item, and in the light of these the item classification is modified (by, for example, rejecting items with poor swags, or classifying items at a different level). The scoring rules might also be modified at this stage (if, for example, too many items appear to be discriminating at the wrong level, or if a high proportion of children's performances do not scale -in the sense of children who have been assigned to a given level failing at a lower level). The children's levels are then reassigned, new swags are drawn and the cycle repeated until some kind of equilibrium is reached.

The method was tried on the Algebra test with a reasonable amount of success: for example, one scale derived by the method, PIG 3, which consisted of 32 items classified into 4 levels, included 23 of the items used for the final Algebra levels. However, it was felt that the method was not entirely satisfactory. First, the method is circular, since all, or most of the items being investigated are used to determine the levels of the children which in turn are used to re-classify the items. This circularity is not serious when a large number of items is involved at each level, but this was not the case for the Algebra test. Second, though some poorly discriminating items can easily be identified by inspection (eg item 4iv below), it is difficult to find precise criteria for interpreting swags. For a sharply discriminating item, the middle portion of curve should be "steep" and its end-point "high". However, it is not clear how these features can be expressed as a single quantity, or whether they are even adequate. For example, though in terms of these features 9iv is certainly a better discriminator than 4iv, it might be argued that item 23 is at least as good, since about five times as many children at level 4 pass the item than at level 3.

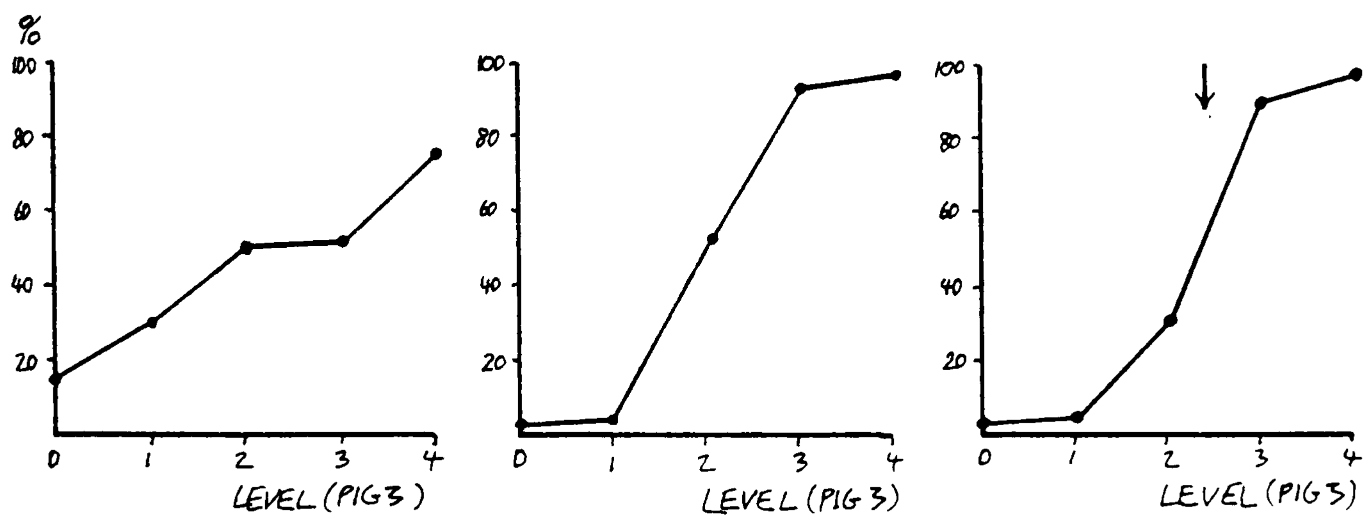
Fig 8.3 Characteristic curve (swag) for items 4iv, 9iv and 23.



Because of these difficulties, a method of analysis was eventually adopted based on the correlation coefficient phi. In general the values of phi supported the intuitive judgements derived from the swags. For example, the mean value of phi for 9iv with all the other items on the Algebra test was substantially higher than the mean for 4iv (see below). On the other hand, the mean for 15ii was even higher, presumably because it discriminates adequately over a wider range, though its discrimination is not as sharp. (Appendix 8 shows the swags for all the Algebra items, under scale PIG 3, and all the mean phi values.)

Fig 8.4 Characteristic curves (swags) for items with increasingly high discrimination, and mean phi for each item.

Item 4iv	Item 15ii	Item 9iv
mean phi: 0.29	mean phi: 0.57	mean phi: 0.44



Chapter 9 MEASURES OF ASSOCIATION BETWEEN ITEMS: THE PHI COEFFICIENT

The various wrong answers that children gave to the items on the Algebra test played a crucial role in making sense of the way children performed on the test (Chapter 5). However, for purposes of statistical analysis, responses to the items were classified dichotomously, as right or wrong.

This chapter discusses some of the ways in which the degree of association between pairs of dichotomous items might be measured. To illustrate the relationship between such pairs of items it is useful to draw contingency tables, and in this chapter these tables are presented in the form shown below, using the following conventions:

0 represents "wrong response" or "fail",
1 represents "correct response" or "pass";

a, b, c and d represent the percentage of children in the categories pass-fail, pass-pass, fail-fail and fail-pass respectively, for items P and Q, although for ease of exposition these will sometimes be treated as frequencies;

p and q are the facilities of items P and Q;

$p \geqslant q$, ie the easier item is shown on the left of the table.

N represents the number of children attempting each item.

Fig 9.1 General contingency table for pair of dichotomous items.

	1 p	a	b
Item P			
	0 p'	c	d
	N	q'	q
		0	1

Item Q

Regardless of how much easier item P is than item Q, for the items to be closely related to each other d should ideally be zero: for example, if P and Q have the same facility they can be regarded as measuring the "same thing" if the children tested respond to the items in exactly the same way (ie they fail or pass both items, as in the first contingency table below where P and Q both have facilities of 60%); if the items are different in facility they can be regarded as measuring the same thing, or at least measuring parts of the same construct, if none of the children who fail the easier item (item P) pass the harder one (as in the second table below where P and Q have facilities of 90% and 60% respectively).

Fig 9.2 Ideal (d=0) contingency tables for 60% by 60% and 90% by 60% facility item pairs.

P	1	60	0	60
	0	40	40	0
			40	60
			0	1
			Q	

P	1	90	30	60
	0	10	10	0
			40	60
			0	1
			Q	

Thus the simplest wav to measure the degree of association between pairs of dichotomous items is to see how close d is to zero (as occurs, for example, in the ordering method proposed by Bart and Krus,1973). However this does not always lead to an equitable comparison because though d can in theory always be zero its maximum possible value is limited by p' or q (whichever is the smaller): for the first pair of items illustrated above d could have a value of 40 whereas for the second pair d can not be greater than 10. A more satisfactory use of the information contained in a contingency table is to compare the products bc and ad in some way, by for example forming the quotient bc/ad or the difference bc-ad. The product bc can be interpreted as the number (when N=100) of subject-pairs that are ranked concordantly on the two items (ignoring ties), whilst ad gives the number of discordant rankings. A disadvantage of the quotient bc/ad is that it becomes excessively large as d approaches zero, and for this reason it will not be considered further. Measures of association are conventionally scaled to range from -1 to 1 (complete disagreement to complete agreement) and to equal zero when there is complete independence. The difference bc-ad satisfies the latter condition (the items are independent when $b/a = d/c$, or $b/d = a/c$) and its range -ad to bc can be mapped onto -1 to 1 by choosing a suitable denominator. Four such transformations are shown below (which result in the coefficients phi, Y, Q and H_{ij}).

$$\text{phi} = \frac{bc - ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad \text{(eg Guilford,1965)}$$

$$Y = \frac{bc - ad}{bc + ad + 2\sqrt{abcd}} \qquad \text{(Yule,1912)}$$

$$Q = \frac{bc - ad}{bc + ad} \qquad \text{(Yule,1912)}$$

$$H_{ij} = 1 - \frac{(a+b+c+d)d}{qp'} \qquad \text{(Loevinger,1947)}$$

$$= \frac{bc - ad}{(b+d)(c+d)}$$

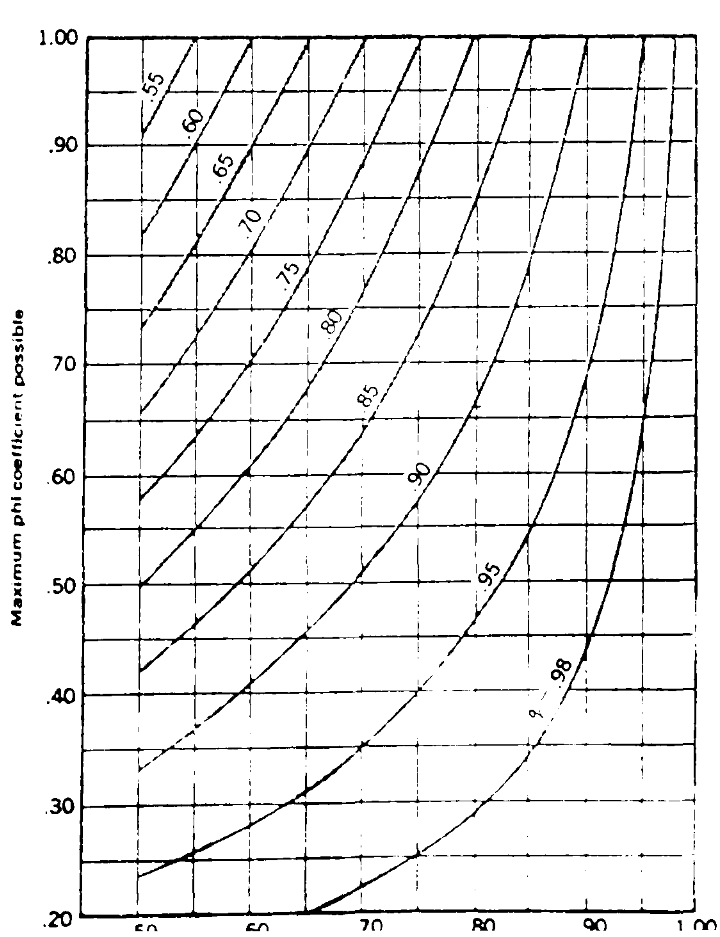
The best known of these coefficients is phi, being infact the product-moment correlation coefficient for dichotomous data, and for this reason phi was initially thought to be the obvious choice for assessing the degree to which pairs of items could be said to be testing the same construct. In the event, phi was chosen but not before the other coefficients listed above had been examined very closely. The examination of these coefficients was prompted by the sudden realisation that the value of phi, though theoretically being able to range between -1 and 1, was infact restricted by the relative facilities of of the items being compared: for example, for the two ideal cases (d=0) displayed earlier, phi has a value of 1 when the facilities of P and Q are both 60% but is only about 0.4 when the facilities are 90% and 60% respectively (whereas Y, Q and H_{ij} equal 1 in both cases). Put in more general terms, when two items are as closely related as possible, ie when d=0, phi can only have a value of 1 when the items have the same facility, whereas Y, Q and H_{ij} will always equal 1.

This discovery seemed to suggest that all that remained to be done was to determine whether one of these latter coefficients behaved in any sense more appropriately than the other two. Unfortunately, when Y, Q and H_{ij} were applied to empirical data it became apparent that they suffered from a complementary problem to that of phi: whilst the value of phi decreased as the difference in facility between item-pairs became more extreme, the values of Y, Q and H_{ij} increased, rather than staying more or less the same. The severity with which this happened varied from one coefficient to another, but given that the Algebra test (and the other CSMS tests) had been deliberately designed to include very easy and very difficult items, the problem was more serious for each of these coefficients than for phi: in effect Y, Q and H_{ij} were saying that the relationship between item-pairs was at its strongest when the information on which to base such a decision was minimal and at its least reliable (for a very easy and a very difficult item, the values of b, c and d are by definition very small, and any fluctuation in their values is likely to have a substantial effect on ad in particular, and hence on Y, Q and H_{ij}, whereas the value of phi, which is already small, will stay small in absolute terms). Expressed another way, Y, Q and H_{ij} were placing the greatest emphasis on the relationships that were the most trivial, ie they emphasised the instances where children who could answer an extremely difficult item also coped with an easy one: in the context of algebra, or school mathematics generally, it would be rare for this not to happen. Thus, after much deliberation, phi was chosen as the most satisfactory measure, in the sense that it did not overestimate the relationship between items, but instead erred on the side of caution.

The graph below (from Guilford,1965,p336) shows the maximum value of phi for different values of p and q, using the formula

$$\text{phi}_{max} = \sqrt{\frac{p'q}{pq'}}$$

Fig 9.3 **Maximum possible phi values for given facilities.**



The restrictions on phi illustrated by the graph can also be seen in the empirical data presented in the table below, which shows values of phi (decimal point omitted) for a group of mid-facility Algebra items against mid- to progressively lower-facility items, obtained on the total 1976 sample (N=2923): as the difference in facility widens, phi decreases. (The phi values for all item pairs are shown in Appendix 9.)

Fig 9.4 Phi values for mid-facility Algebra items against mid to low facility Algebra items.

mid-facility					low-facility	
· 39 35 32 42	15 34 25 30 38	30 31 34 32 28	24 26 28 22 21	24 22 23 21 19	18 20 18 15 11	mid-facility
· 50 42 54	21 48 38 41 46	48 41 44 42 40	33 33 37 28 29	32 29 28 28 26	22 22 20 15 14	
· 69 48	18 63 45 65 45	39 33 40 43 37	35 31 36 34 29	32 33 28 27 27	20 22 20 15 14	
· 41	14 51 39 54 37	35 28 34 35 32	36 27 30 37 23	26 34 26 23 22	18 24 21 16 15	
·	20 49 38 43 55	47 40 52 49 42	38-36 42 31 32	36 30 31 34 32	25 28 25 19 18	

The coefficient Y was devised by Yule (1912) in an attempt to solve the very practical problem of assessing the effectiveness of different vaccination programmes against smallpox. Given hospital records something like the ones below, Yule argued that, at first sight, it would seem reasonable to compare the effects of the vaccination in towns A and B by examining either the difference b/q - a/q' for both towns, or the difference b/p - b/p' for both towns (ie the difference between the proportion of those who lived who were vaccinated and those who died who were vaccinated, or the difference between the proportion of those who were vaccinated who lived and those who were not vaccinated who lived). However, the first difference

Fig 9.5 The effect of vaccination in Towns A and B.

		Town A			Town B	
Vaccinated	60	10	50	75	1	74
Not Vaccinated	40	10	30	25	1	24
		20	80		2	98
		Died	Lived		Died	Lived

makes the programme in town A come out worse ($\frac{50}{80}-\frac{10}{20}=\frac{1}{8} < \frac{74}{98}-\frac{1}{2}=\frac{25}{98}$) than in town B, whilst the second difference makes the programme in town A come out better ($\frac{50}{60}-\frac{30}{40}=\frac{1}{12} > \frac{74}{75}-\frac{24}{25}=\frac{2}{75}$). Yule went on to argue that the reason why these comparisons are not equivalent is that different proportions of people were (randomly) vaccinated in the two towns, and the outbreak of the disease was not of the same virrulence: thus the proportions of those vaccinated and not vaccinated, and of those who lived and died should first be made equal, which, according to Yule, may be done by choosing suitable scalar multipliers for each row and column (with the result that the marginals all have a value of 50).

The coefficient Y gives the value of both $b/q - a/q'$ and $b/p - b/p'$ after these transformations have been performed, and it has the remarkable property that it is unaffected by any of the transformations. Further, Y is equal to phi when the marginals are all 50 to start with. However, while the transformations may be valid in the context where Yule applied them, it is not clear that this is also the case for items. Consider, for example, the transformation illustrated below (where the second row of the contingeny table has been doubled): in theory it would be possible to find another 25 children who performed in the same way as the original 25 who failed item P; however, assuming that such children are of lower ability than those who succeed on item P, the ability distribution of the new sample would be substantially altered, which suggests that Y might be describing the relationship between items in circumstances (for samples) that are atypical.

Fig 9.6 Transformation on one row of contingency table.

Item P	1	75	25	50	$\longrightarrow$	1	75	25	50
	0	25	15	10		0	50	30	20

In practice, it must be said that the values of Y were reasonably uniform. Thus in the adjacent table. which shows the values of Y for the easiest five items on Algebra test against items of decreasing facility, the increase in Y is very gradual, except at the very extreme, where in several cases Y shoots up to 1. (Again, these values are for the 1976 total sample, N=2923.) Because of the generally small sampling variation in Y, Shayer (1978a) chose Y in preference to phi for a factor analytic study of the CSMS science tasks. For extreme differences in facility, his solution was to argue that

> "There is no way out of the necessity to inspect deliberately the correlation matrix and eliminate those variables (items) which give Y values very close to or equal to 1, on the grounds that there is too little information for the coefficient to estimate the correlation"(ibid,p163).

Arguably this is as good a solution as accomodating factor analysis to the ideosyncrasies of phi (see Chapter 10). However, it still leaves doubts about the "typicalness" of the relationship described by Y that was referred to above, and in

high-facility

·					high-facility
51	·				
70	41	·			
54	51	52	·		
46	34	47	46	·	
51	38	51	49	86	
40	40	39	37	31	
54	41	50	46	37	
36	44	33	32	35	
45	43	41	39	29	
33	46	44	45	39	
34	42	42	37	30	
59	43	54	43	38	
38	53	40	49	35	
54	39	38	44	35	
26	36	39	31	25	
52	34	73	44	36	
48	42	53	48	48	
58	35	76	45	36	
62	52	43	52	40	
40	36	43	49	36	
62	58	58	63	49	
57	48	60	56	44	
54	42	58	54	39	
60	58	60	58	46	
30	29	37	28	24	
54	43	53	54	46	
49	42	46	41	35	
54	40	51	53	41	
46	61	49	61	48	
65	49	49	54	46	
41	46	44	63	36	
51	43	60	54	53	
34	55	46	51	41	
66	57	47	59	43	
49	40	45	51	49	
47	51	57	44	38	
59	65	53	74	43	
50	48	51	49	44	
67	39	62	44	44	
54	70	59	56	61	
64	70	53	69	47	
60	55	61	65	59	
57	100	69	100	67	
55	100	67	100	74	
100	100	100	49	100	
100	100	62	55	100	
44	100	58	100	100	low-facility

Fig 9.7 Y values for high-facility Algebra items against high to low-facility Algebra items.

particular about the feeling that Y overestimates the relationship (even if to a lesser extent than Q and H_{ij}), which is reinforced by the very high communalities that occur when Y is used in a factor analysis (even after items with extreme facilities have been eliminated).

Yule's Q (named after Quetelet) is closely related to Y, and has the same property that it remains unaffected by scalar multiplication of the contingency table rows and columns. However, its value is consistenly greater than Y and the problem of overestimation is therefore more acute.

Q can readily be interpreted in terms of the number (or proportion) of concordant (bc) and discordant (ad) subject-pairs. On the other hand, Loevinger (1947) defines her "homogeneity" coefficient H_{ij} in terms of individuals, namely those (d) who pass the harder item and fail the easier, each of whom "indicates a discrepancy in classification according to the two items"(ibid,p36). The maximum possible number of such discrepancies is equal to the minimum of p' and q; however, rather than comparing the observed d to this value, Loevinger uses the chance expectancy of d, so that when d has this value H_{ij} is zero. The chance expectancy is given by $q \times p'/N$ (where q and p' are frequencies rather than percentages) and H_{ij} is then defined as

$$H_{ij} = 1 - \frac{d}{qp'/N}$$

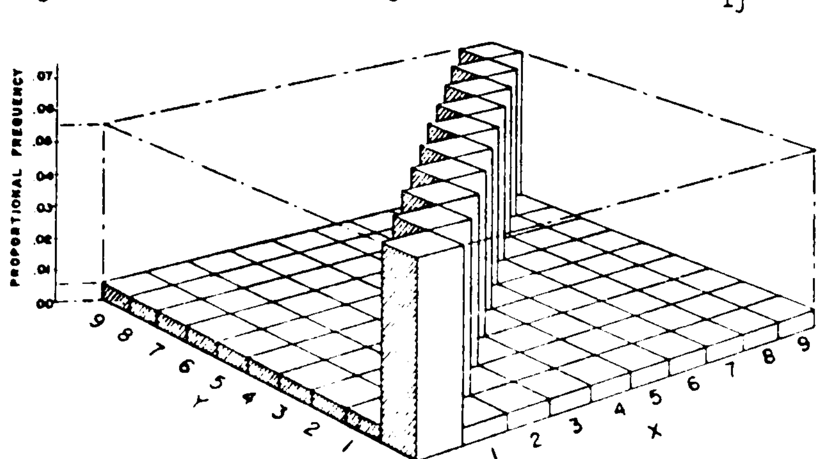
$$= 1 - \frac{dN}{qp'}$$

$$= \frac{bc - ad}{(b+d)(c+d)} .$$

(where a,b,c and d may now be thought of as frequencies or percentages)

Loevinger (1948) also points out that H_{ij} is equal to the ratio phi / phi_{max}. Both these interpretations are highly intuitable, which makes H_{ij} very attractive. Unfortunately, however, its sampling variation is even greater than Q, as well as Y. Carroll (1961) is highly critical of H_{ij}. He sets two conditions on a measure of association based on a 2 x 2 table, namely that its value should equal the Pearsonian correlation computed for the underlying correlation surface and that its value must be independent of the two dichotomisation points chosen to yield frequencies in the 2 x 2 table. As far as H_{ij} is concerned, these conditions are only met for a bivariate rectangular distribution like the one illustrated below (from Carroll,1961,p363). (It should be said that for a normal bivariate distribution, phi does not meet these conditions either, in that it underestimates Pearson's r, as well as being dependent on the dichotomisation points. Carroll favours the tetrachoric correlation r_t, and it is perhaps a pity that it was not investigated here, along with Y, Q and H_{ij}. However, Shayer (1978a) argues that r_t is particularly susceptible to inconsitent variation and evidence presented by Carroll himself seems to support this, as does the relationship between the unstable quotient bc/ad and the estimates of r_t tabulated by Davidoff and Goheen (1953).)

Fig 9.8 Bivariate rectangular distribution for H_{ij}



Illustrating the Type of Bivariate Rectangular Distribution for which ϕ/ϕ_{max} is Constant and Equal to the Pearsonian r for the Distribution as a Whole (r = 5

Kaltenhauser and Lee (1976), using simulated values, found that the sample variance of phi was much smaller than those of phi/phi_{max} (H_{ij}) and the tetrachoric; for medium-sized values in particular, the latter coefficients had standard deviations 50% to 100% larger than phi. A similar picture emerges from a factoranalytic study by Comrey and Levonian (1958) who found excessively high communalities using phi/phi_{max} and r_t, but not, in general, with phi. Shayer (1978b) undertook an empirical comparison of phi, Y, H_{ij} and r_t, using tables of normal bivariate distributions (N=1000) in McNemar (1969, pages 147 and 148). For r=0.5 he split the table dichotomously in 8 different ways, and found the following means and standard deviations for the coefficients:

Fig 9.9 Empirical values of phi, Y, H_{ij} and Q.

	phi	Y	H_{ij}	r_t
mean	0.25	0.42	0.54	0.62
standard deviation	0.061	0.055	0.232	0.067

In this case, the sampling variation was about the same for phi, Y and r_t but H_{ij} varied drastically, whilst phi in particular underestimated the true correlation. The table below shows the values of phi, Y, Q and H_{ij} between the Algebra PIG 8 levels P1, P2, P3 and P4 split dichotomously, ie being treated as items, for the total 1976 sample (N=2923). As expected, phi decreases as the facility difference between the levels increases, whilst Y and especially Q and H_{ij} grow larger.

Fig 9.10 Values of phi, Y, H_{ij} and Q for Algebra PIG 8 levels.

PHI	P1	P2	P3	P4
P1	·27	·24	·18	·12
P2		·32	·28	·19
P3			·24	·28
P4				·32

Y	P1	P2	P3	P4
	·39	·34	·35	·44
		·34	·34	·40
			·36	·41
				·47

Q	P1	P2	P3	P4
	·68	·61	·62	·75
		·60	·61	·69
			·64	·70
				·77

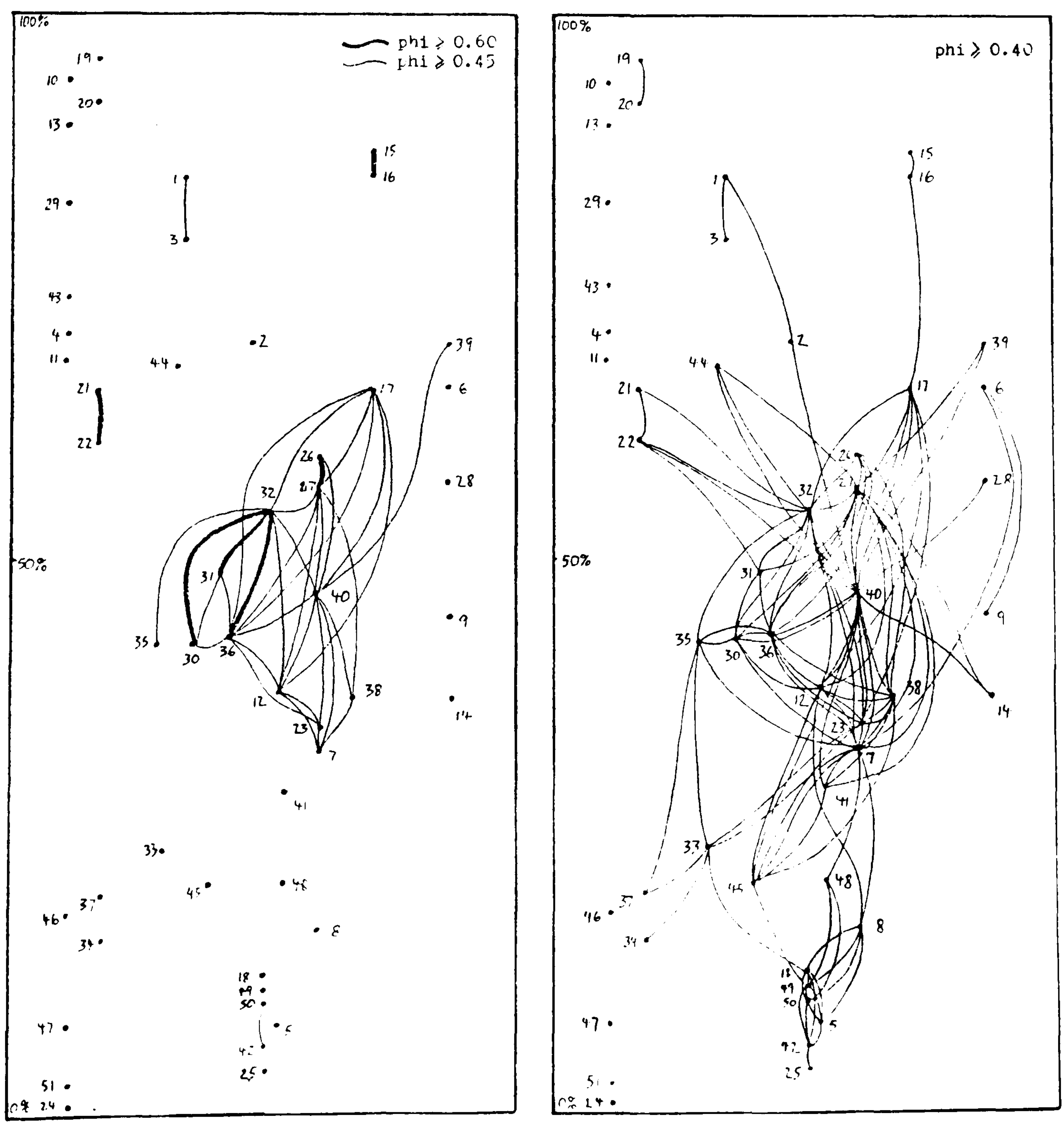
H	P1	P2	P3	P4
	·27	·45	·64	·80
		·32	·53	·71
			·34	·55
				·32

Chapter 10 SPIDER DIAGRAMS AND FACTOR ANALYSIS

Spider Diagrams

The phi correlation matrix shown in Appendix 9 was used to generate "spider diagrams" such as the two shown below. In these diagrams, the vertical axis represents facility (for the total 1976 sample, N=2923) whilst the horizontal axis is simply used to separate the items; in the first diagram, arcs have been drawn between all pairs of Algebra items for which the value of

Fig 10.1 Spider diagrams (phi>0.45 and phi>0.40) for Algebra items.



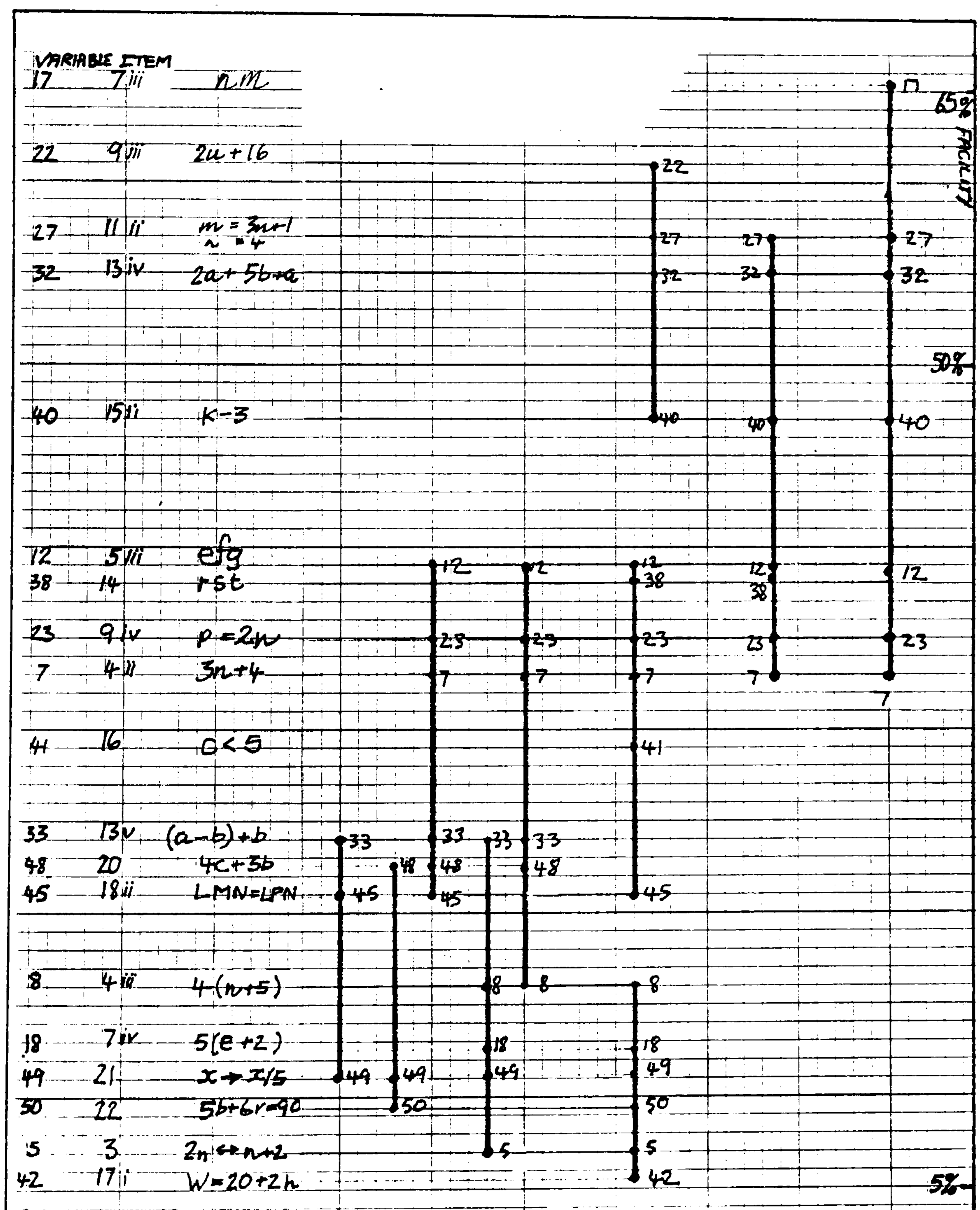
phi is greater or equal to 0.45 and 0.60, whilst in the second diagram the limit on phi has been lowered to 0.40. The items themselves have been named according to their variable numbers (which is how they were coded for the computer). The corresponding item numbers are listed in Appendix 10.

Such diagrams were drawn for all the CSMS mathematics tests, and they proved to be extremely useful for acquiring an overall impression of the structure of the test, and for making an initial identification of the items that seemed to correlate best or worst with the test as a whole (eg variables 27, 32, 40, 12, and 46, 47, 51, 24 respectively).

An interesting feature of the spider diagrams generally, and which is shown very clearly by the first diagram above, is that the phi values tended to be highest not just for items of similar facility (the restricted-range problem) but for items at about the middle of the facility range; presumably this occurred because such items provide more information about the underlying relationship between them, so that the value of phi gets closer to the "true" correlation, ie to Pearson's r. As for items at the extremes of the facility range, it was generally the case that easy items correlated less well with each other than difficult items (as is shown especially by the second diagram above); since none of the items were multiple-choice, a likely explanation for this is that the easy items contained a greater proportion of "errors", in the sense that a greater proportion of children passing the difficult items failed on an easy one (through carelessness, say) than children failing the easy items passed a difficult one. Another feature that the spider diagrams show very clearly is that items tend to correlate more highly with other items from the same question: for example, the highest correlations (phi≥ 0.60) on the first diagram above are between variables 15 and 16 (items 7i and 7ii), 21 and 22 (items 9ii and 9iii), 26 and 27 (items 11i and 11ii), and 32 with 30, 31 and 36 (which are all parts of Question 13). Similarly variable 9 (item 4iv), which turned out to be one of the least consistent items on the test and has no phi values equal to or above 0.45, nonetheless has one link appearing on the second diagram, to variable 6 (item 4i) which in turn links up to variable 7 (item 4ii).

A way of trying to overcome the problem of phi's restricted range is to assume that the relationship expressed by phi is transitive. Taking the first spider diagram above, it can then be argued, for example, that variable 27 is related to 7 because of the link between 27 and 40, and 40 and 7. Ofcourse, such an inference has to be treated catiously: the fact that a "has something in common with" b, and b with c, does not necessarily mean that a and c are similar in any way; on the other hand this particular inference is supported by the second diagram, which shows that there is a link between variable 27 and 40 when the value of phi is lowered. An attempt was made to apply this argument more methodically, by drawing modified spider diagrams like the one below, where each chain of items represents a lattice for phi greater or equal to 0.38 (ie the value of phi between each pair of items in a chain is at least 0.38). There is a substantial overlap between one lattice and the next, which suggests that the 19 items shown on the diagram are all (to some degree..) measuring the same construct. In the event, these items were all later selected to form the Algebra levels, but it is noticable that none of the easiest Algebra items are represented on the diagram, so that if a "starting off" level was to be constructed for the Algebra test a lower criterion (in terms of the value of phi) would have to be adopted for these items.

Fig 10.2 Lattice diagram (phi>0.38) for Algebra items.



One weakness of spider diagrams, even of the lattice version, is that it is difficult to describe the degree to which items appear to be testing the same construct, other than on an intuitive level. Another problem is the choice of a lower limit for phi, and perhaps more to the point the consequence of having to make such a choice at all. There are no objective criteria for deciding that a given level of phi is "good": the best that can be hoped for is a consensus of opinion, once a sufficient number of people have considered the problem; even tests of significance for determining that a particular value of phi is "bad" are irrelevant, given the large number of children tested (despite the very simple relationship between phi and chi squared). The fact that a limit for phi has to be chosen when drawing a spider diagram also means that many of the genuine (albeit weaker) relationships between the items are completely ignored. The peculiar choice of the value 0.38 in the above diagram (it had originally been intended to choose a limit of 0.40) is testimony to the frustration that this causes. Because of these various shortcomings, it was decided to supplement the use of spider diagrams with factor analysis, in the hope that this would provide a more systematic and comprehensive way of determining the strength of the relationship between the items. By and large this was the case, though the ideosyncrasies of phi still meant that the outcome had to be interpreted with care.

Factor Analysis

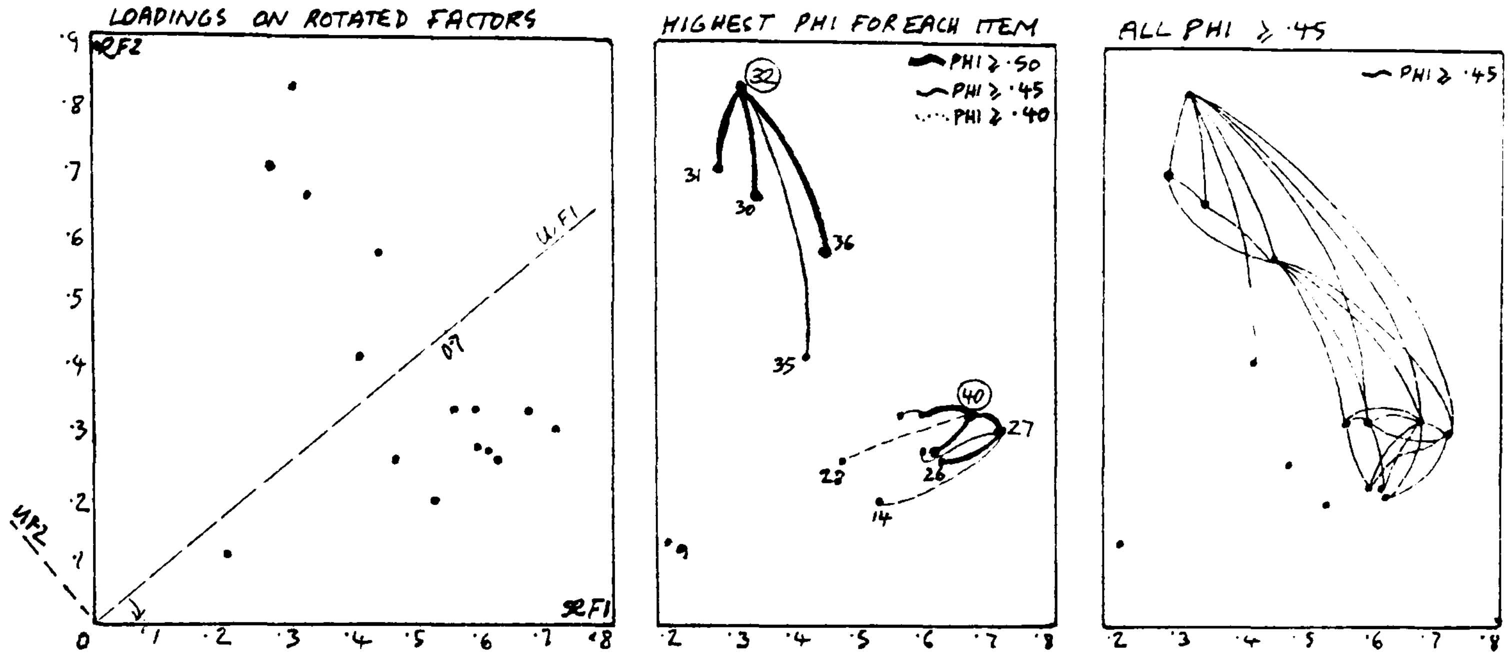
The method adopted was the SPSS (1975) principal factor programme PA2 (although principal component analysis and other forms of factor analysis such as image and alpha were also tried, but they appeared to make little difference to the picture that emerged).

Given that the first unrotated factor expressed substantially more of the test's variance than any other factor (which it always did), it was hoped that by simply examining the item loadings on this factor it would be possible to identify immediately the items that correlated best with the test as a whole. However, a number of influences complicated this apparently simply procedure.

The first and most obvious of these is that the nature of the first factor, and therefore the loading of items on it, can be severely distorted by a preponderence of items from the same question. On the Algebra test the majority of questions have at most four parts, but one has nine parts and it was noticed that when all the items were subjected to factor analysis the first unrotated factor was heavily pulled towards these items. This meant that instead of the first factor being an immediate and "neutral" indicator of the basic nature of the test, a decision first had to be made about what selection of items could be regarded as representative of the test. In the end, the rule was made not to allow two items from the same question if these were close in facility and to reject the item with the lower loading.

A related influence, which has already been mentioned in the context of the spider diagrams, was the tendency for items from the same question to correlate more highly with each other than with other items, with the result that they tended to generate their own factor. This is illustrated by the first of the three diagrams below, which shows loadings on the first two factors (after rotation) obtained by factoring just 15 mid-facility items from the Algebra test, 5 of which came from Question 13. (As with all the analyses, the total 1976 sample was used, with N=2923.) The result is to partition these items into two groups, or

Fig 10.3 Loadings on RF1 and RF2, highest phi and all phi ≥ 0.45 for mid-facility Algebra items.



clusters, depending on whether the items are part of Question 13. In a sense these clusters are real enough, since there is a strong link between the items within the clusters, as can be seen from the second diagram above. However, the third diagram shows that the clusters are also strongly linked to each other, and indeed the variance that the items have in common (ie on the first unrotated factor, UF1) is about 8 times greater than that due to their specific differences (the variance on the second unrotated factor, UF2). Such specific factors are essentially trivial, and it was decided to take note only of factors that had substantial loadings (above 0.3) on items from different questions, and effectively to eliminate question-factors by selecting only a limited number of items from each question in accordance with the rule mentioned above. The alternative is to attach significance to the specific features of every question on the test, and every other question that might have been written. A similar observation has been made by Furneaux and Rees (1978,p507) who speak of "simple content factors".
For the 15-item analysis, the table below shows the percent of variance and the corresponding eigenvalues for each factor before communality estimates have been made (at which stage the

factors are strictly components), and for the two factors with eigenvalue greater than 1, after the communality has been determined by iteration.

Fig 10.4 Eigenvalues and percent of variance of components, for mid-facility Algebra items.

Factor (Component)	Eigen-value	Percent of total variance
1	6.52	43.5
2	1.14	7.6
3	.96	6.4
4	.87	5.8
5	.76	5.1
6	.68	4.5
7	.62	4.2
8	.57	3.8
9	.52	3.4
10	.49	3.3
11	.46	3.1
12	.45	3.0
13	.43	2.9
14	.27	1.8
15	.24	1.6

Factor	Eigen-value	Percent of variance of first two factors
UF1	6.01	89.6
UF2	.70	10.4

When the analysis of the mid-facility items was repeated, but with 3 of the 5 parts of Question 13 removed (ie 13iii and 13iv remaining), only one of the initial factors (components) had an eigenvalue greater than 1 (5.27, which represents 43.9% of the total variance, compared to 43.5% above).

A third complication is the tendency for factor analysis to generate facility factors, especially if the items span a wide facility range. This seems to have been first noticed by Ferguson (1941), and stems from the restricted-range property of phi. An early attempt to overcome (or rather to circumvent) the problem is illustrated by the table below: items were classified into overlapping groups of restricted facility and factored separately, and the attempt was then made to match the loadings from one set of factors with the overlapping loadings from another set. The table shows the loadings (decimal point omitted) for the items (variables) ordered by facility on five sets of rotated factors. A comparison of the loadings suggests, for example, that factor $F4_1$ (called E_1) is the same as $F1_2$ (E_2), since they have very similar loadings on variables 4 (.49 and .49), 11 (.51 and .48), 44 (.51 and .52), 21 (.26 and .25), etc. Unfortunately the method was not entirely

Fig 10.5 Loadings on rotated factors for factor analyses of items of similar facility.

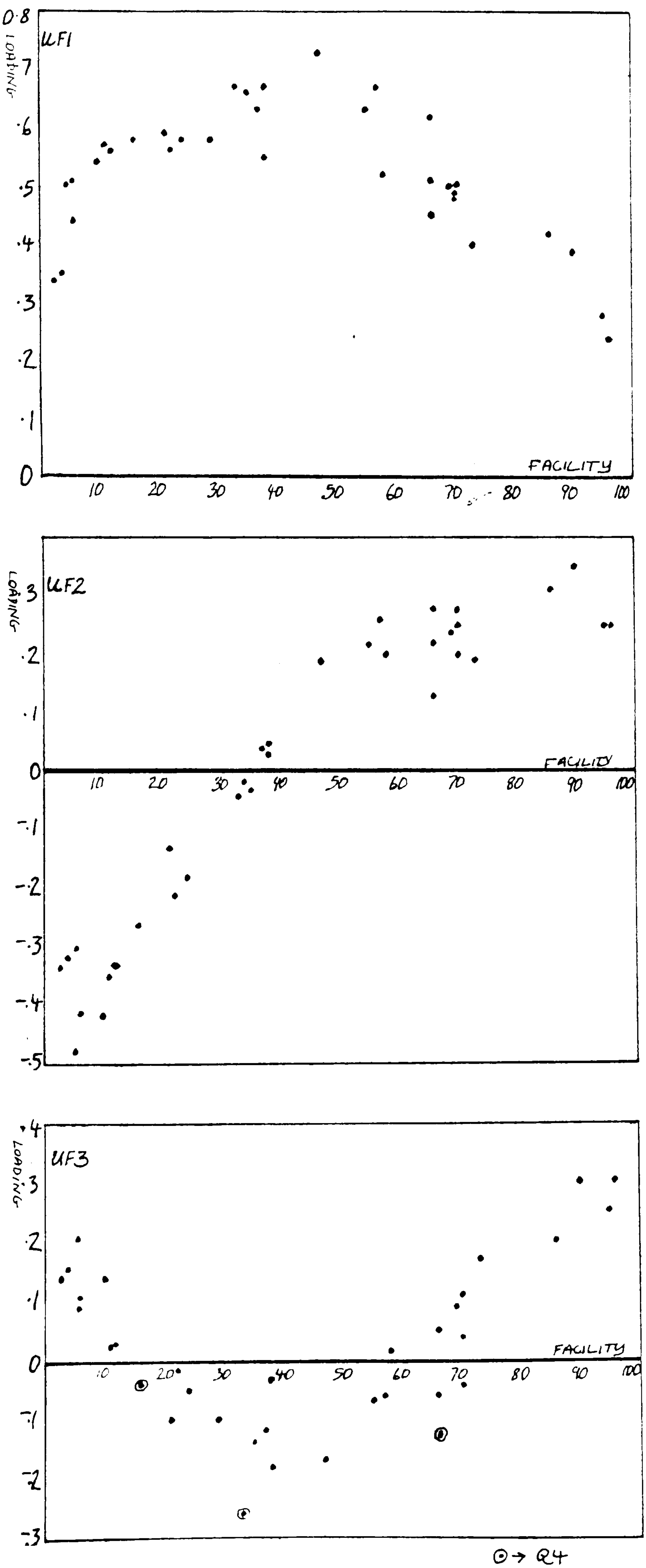
ROTATED FACTORS

	VARIABLES	$F1_1$	$F2_1$	$F3_1$	$F4_1$	$F5_1$	$F1_2$	$F2_2$	$F1_3$	$F2_3$	$F1_4$	$F2_4$	$F3_4$	$F4_4$	$F1_5$	$F2_5$	$F3_5$
easy	19	09	06	08	13	62											
	10	04	01	18	32	23											
	20	15	22	12	23	54											
	13	17	10	12	40	32											
	15	83	09	12	21	15											
	16	83	09	14	27	19											
	1	08	05	67	19	14											
	29	14	15	12	35	31											
	3	11	10	61	18	11											
	43	08	09	09	40	22											
	4	16	14	15	49	08	49	21									
	2	06	16	50	37	05	46	23									
	39	13	10	18	46	23	51	18									
	11	11	14	12	51	10	48	21									
	44	10	12	11	51	15	52	19									
	6	06	13	18	40	04	39	18									
	21	10	81	12	26	17	25	85									
	17	22	18	20	54	09	62	23									
	22	09	80	15	27	15	28	81									
	26						69	14	62	25	29	22	15	72			
	28						46	30	46	25	39	24	10	24			
	27						79	16	71	30	36	26	19	76			
	32						59	33	31	83	27	81	18	21			
	31								27	70	20	67	25	16			
	40								67	33	59	30	17	30			
	9								22	14	08	06	29	14			
	36								44	58	36	52	28	22			
	35								44	41	26	32	47	16			
	30								33	66	28	60	28	14			
	12								59	33	55	27	25	19			
	38								59	27	50	20	27	26			
	14								52	19	45	15	18	23			
	23								60	27	57	20	25	14			
	7								55	33	56	23	32	12			
	41										47	20	26	17	53	08	19
	33										36	20	52	07	61	19	14
	48										46	12	32	09	45	29	24
	45										51	18	28	12	55	17	19
	37										19	24	57	06	51	22	10
	46										37	13	21	12	24	06	70
	8										37	11	51	10	53	27	22
	34										28	20	47	06	56	19	09
	18														53	29	21
	49														51	33	25
	50														40	50	24
	47														24	24	69
	5														46	38	25
	42														40	57	17
	25														16	58	10
	51														27	34	18
hard	24														07	34	01
		A_1	D_1	B_1	E_1		E_2	D_2	E_3	F_3	E_4	F_4	G_4	H_4	G_5	F_5	H_5

satisfactory, in part, perhaps, because at this stage the ideas about question-factors had not been worked out and these factors were confusing the picture: in particular, it was not possible to find a clear factor running through all 5 groups, (although E came close to doing so, with the exception of the hardest group where G_5 had some loadings in common with E_4 but at least as many with G_4).

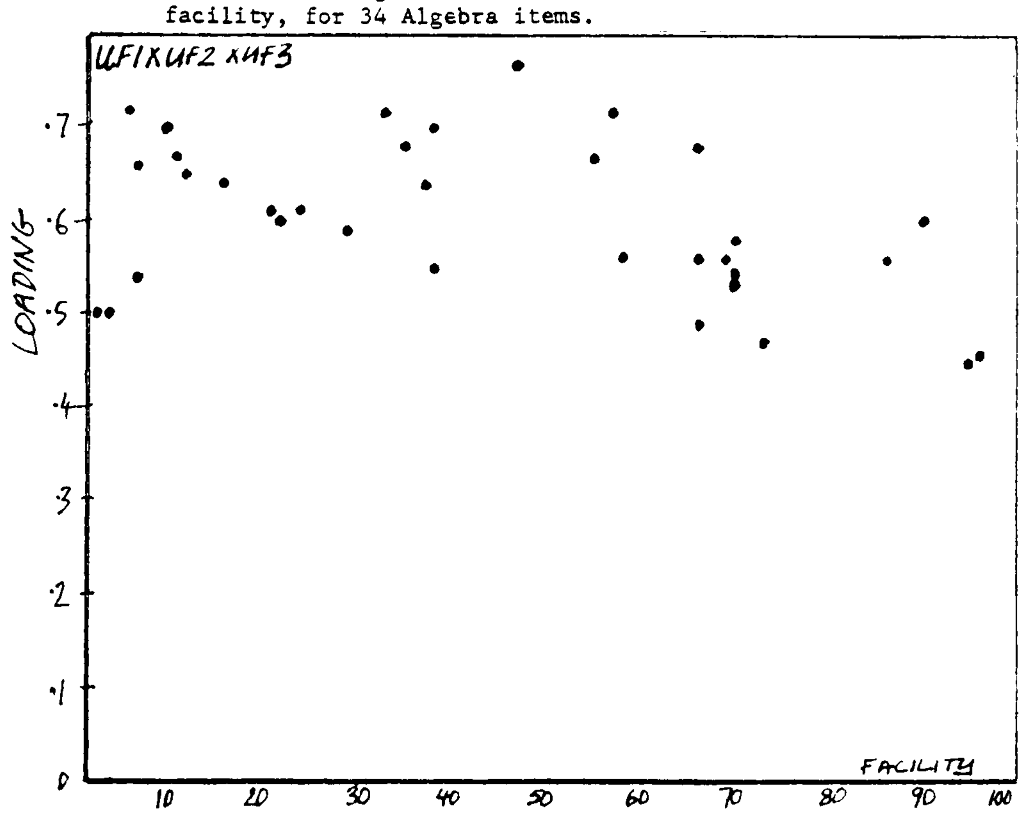
It was later decided to tackle the problem of facility factors head on, by analysing the items in one group, regardless of facility (but after first removing "mates" from the same question, in accordance with the previously mentioned rule). In one such analysis 34 of the 51 items on the Algebra were used, which resulted in 3 factors whose eigenvalues were initially greater than 1. However, the loadings on each of these were strongly influenced by facility, which can be seen very clearly from the graphs below, where the loadings on these three (unrotated) factors have been plotted against facility. Thus it can be argued that the emergence of 3 substantial factors (factors with eigenvalues greater than 1) rather than a single general factor is purely an artefact of phi. Also, there appear to be no discernable factors superimposed on UF1, UF2 and UF3 (which happened for example on the CSMS Graphs test discussed below), other than a possible slight relationship between the three circled items on the UF3 graph (which are all parts of Question 4, but were not removed because it was thought they were sufficiently far apart in facility). Thus it can be further argued that the overall nature of the test is more accurately represented by the combination of these three factors than by the first unrotated factor alone. This can be done by taking

Fig 10.6 Loadings against facility for first three unrotated factors for 34 Algebra items.



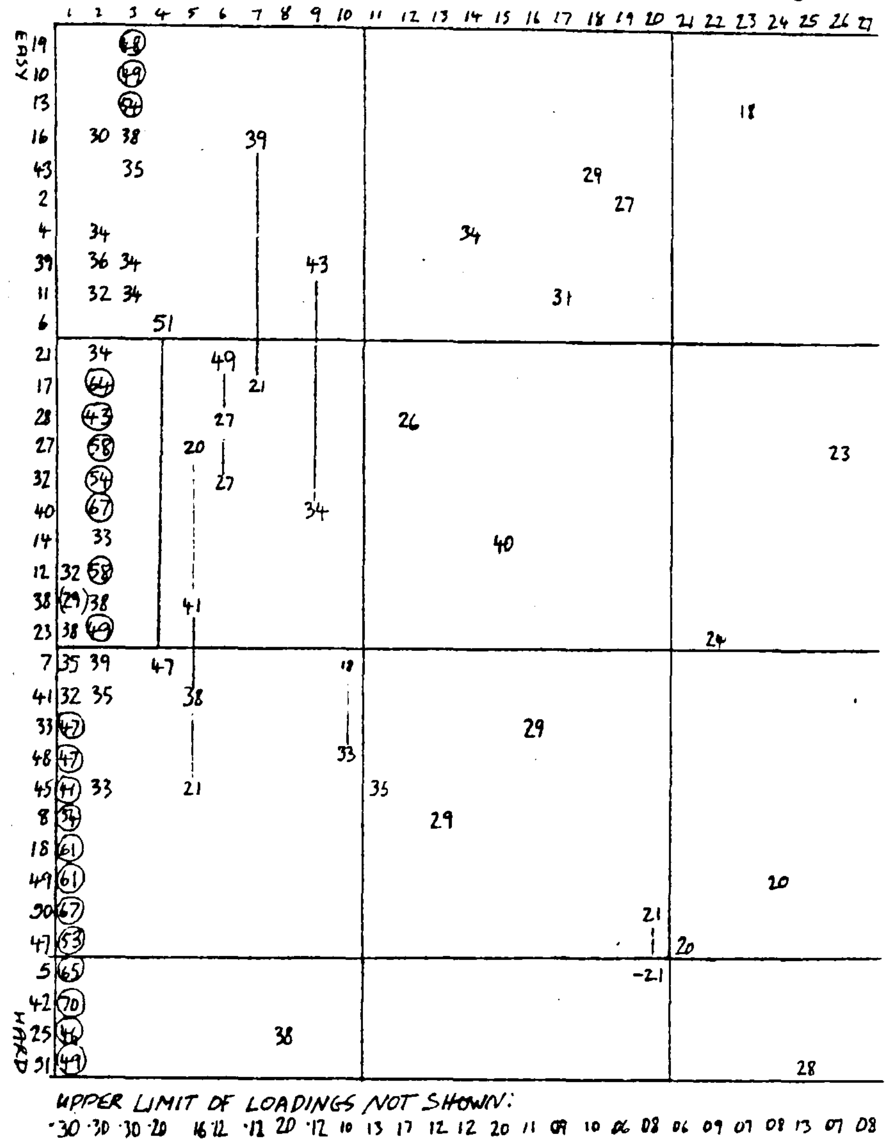
the square root of the sum of the squares of the separate loadings, the result of which is shown below, with the (combined) loadings again plotted against facility. Apart from the tendency for items with high facilities to correlate less well with each other, which has been discussed in the context of the spider diagrams and is not a consequence of the restricted range of phi, there appears to be little or no relationship between the combined loadings and facility.

Fig 10.7 Combined loadings on first three unrotated factors against facility, for 34 Algebra items.



Confirmation that the test can be regarded as one-dimensional comes from the table below, which shows loadings on the factors after rotation. (30 of the initial 34 factors were extracted, even though only the first three had eigenvalues greater than 1.) The items are ordered according to facility, and it can be seen very clearly that the first three rotated factors are purely facility factors. Of the remaining factors, only the fifth shows more than one loading above 0.3 for items not from the same question (variables 38 and 41).

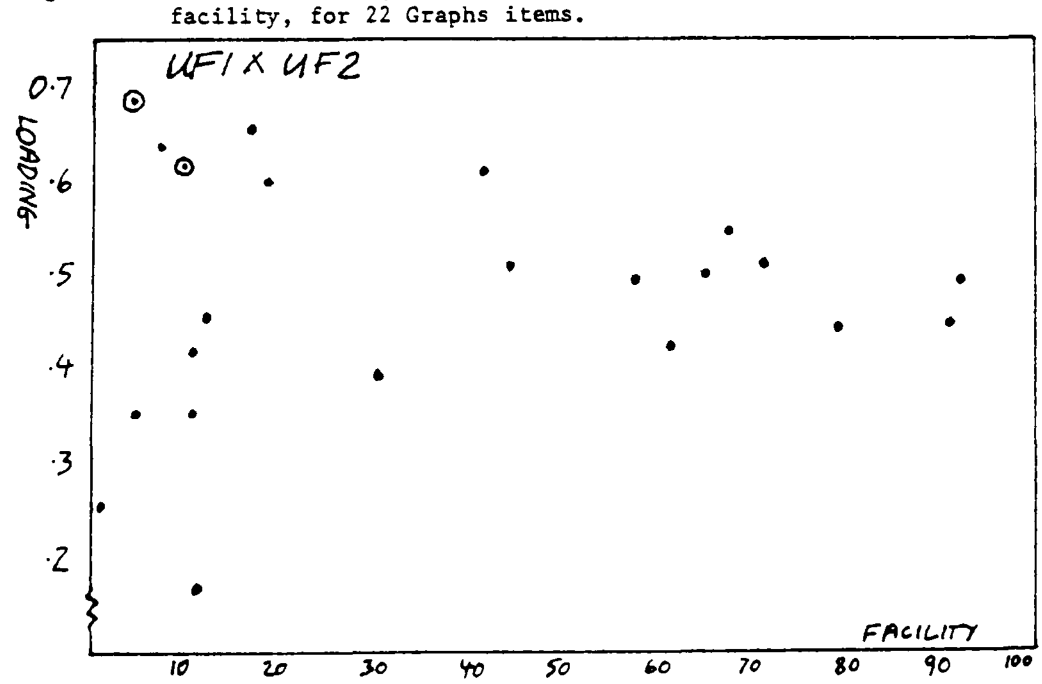
Fig 10.8 Loadings on rotated factors (NF=30) for 34 Algebra items.



A factor analysis was also tried on the CSMS Graphs test, which had been devised by Kerslake (1977). Here, after mates had been removed, four factors were generated with eigenvalues greater than 1. When item loadings for the first two unrotated factors were plotted against facility, similar shaped graphs were obtained as for the Algebra analysis. For the third factor, however, the typical U-shape appeared to be distorted by the loadings of two items that were not from the same question but which both tested the notion of continuity. After rotation, three clear facility factors emerged, as on the Algebra test, but with the difference that the third of these also strongly identified the two continuity items, whose loadings on this factor were both above 0.6. The remaining factor turned out to be trivial, with only one loading above 0.3.

To obtain an estimate of the items' loadings on the test as a whole, it was decided to combine only the first two unrotated factors, since using the third factor would give an overestimate for the continuity items. The combined loadings are given in the graph below, which shows that the continuity items (circled) have a substantial amount of variance in common with the other items on the test, even though in part they seem to be measuring something different. The original analysis of the test based on spider diagrams put these continuity items (and their mates) into a separate cluster, which appeared to be too distinct to be included in the core of items used to define the Graphs levels. However, in the light of the factor analvsis this decision may have been mistaken.

Fig 10.9 Combined loadings on first two unrotated factors against facility, for 22 Graphs items.



Following Shayer's (1978a) example, a factor analysis was also tried on the Algebra test using Y. When all the items were used, the communalities of several items exceeded 1, as Shayer had predicted. Subsequently this was avoided by removing 14 of the easier (and least interesting) items. However, the resulting communalities were still much higher than those obtained using phi, with a bias towards the hardest and the remaining easy items. Thus for example the mid-facility item 15ii (variable 40), which with mates removed had the highest communality of all the items using phi, and with mates was ranked 7th, was only ranked 16th using Y. Further, item 4iv (09), which even with the help of mates was the 10th worst item using phi, was now ranked 24th worst, out of only 37 items. Thus it appeared that Y was not only overestimating the strength of the relationship between the items, but was distorting their rank order and in particular giving more weight to items that on the spider diagrams and the factor analysis using phi had come out very badly.

In summary, the method used to select the Algebra items that correlated best with each other and with the test as a whole, consisted of inspecting spider diagrams supplemented by factor analysis, both based on phi. With the spider diagrams the problem of phi's restricted range was tackled by constructing overlapping lattices, whilst on the factor analysis (after first removing items from the same question with similar facility) one general factor was produced by combining the factors with eigenvalues greater than 1, after determining by inspection that they could be classed as facility factors. Items with the highest (combined) loading on this general factor were then selected to form the levels (though a less strict criterion was used with the easiest items in order to produce a starting-off level, and mates were reintroduced in the light of their performance on previous factor analyses, or in the light of the spider diagrams). The table below shows the mean and median of the phi values between all item pairs within each resulting level.

Fig 10.10 Mean and median phi within each Algebra level.

Level	Mean phi	Median phi
1	0.28	0.28
2	0.42	0.38
3	0.44	0.43
4	0.40	0.40

Chapter 11 GUTTMAN SCALOGRAM ANALYSIS

Guttman scalogram analysis was used, in an informal way, at a very early stage of the research, to assess the consistency of the draft Algebra items (eg Fig 2.12 in Chapter 2, page 31). Later, by means of the SPSS (1975) programme GUTTMAN SCALE, scalogram analysis was used extensively, and intitially with great enthusiasm, on all the CSMS mathematics tests to assess the consistency of the levels.

This form of analysis is concerned with the same problem as that raised in Chapter 9 in relation to phi, Y, Q and H_{ij}, namely, do children who pass the harder of a pair of items also pass the easier one; however, instead of comparing pairs of items, scalogram analysis examines the whole set of items in a test, in order to determine how close the items are to a perfect, (cummulative) "scale" (Guttman,1944), in the sense that subjects who pass one item pass all the other items in the set of equal or higher facility. When this happens, a subject's score on the items uniquely determines the items passed, and the matrix of subjects against items, ordered by score and facility, has what Guttman calls a simplex structure, as in the table below:

Fig 11.1 Scalogram for items forming a perfect scale.

		low facility				ITEMS				high facility
		1	2	3	4	5	6	7	8	9 . . . .
high score	a	1	1	1	1	1	1	1	1	1 . . . .
	b	0	1	1	1	1	1	1	1	1 . . . .
	c	0	0	1	1	1	1	1	1	1 . . . .
SUBJECTS	d	0	0	1	1	1	1	1	1	1 . . . .
	e	0	0	0	1	1	1	1	1	1 . . . .
low	f	0	0	0	0	0	1	1	1	1 . . . .
score	.	.	.	.	.	.	.	.	.	. . . . .

On the CSMS mathematics tests, scalogram analysis was used to see whether the levels formed a scale, rather than the items themselves, ie whether children who passed one level (succeeded on at least two-thirds of one sub-set of items) passed all the lower levels (succeeded on at least two-thirds of the items in each of the corresponding sub-sets of easier items). The reasons for examining the levels in this way were twofold. One was simply to answer the practical question of whether the levels provided a coherent way of describing children's performance on the tests: there would be little point in assigning a group of children to level 3, say, of a given test if a substantial number failed levels 1 or 2. In the event, the proportion of children classified as "error-types" was generally small, the largest being about 8% on the CSMS Reflection and Rotation test devised by Küchemann (1980). This use of scalogram analysis seems to be perfectly valid, but the other reason for using it is more contentious: it is claimed, for example by Loevinger (1948), that scalogram analysis can be used to ascertain how close a test comes to being "uni-dimensional" or "homogeneous", ie how near the items are to testing a single construct, and so it was thought that this form of analysis would provide a measure of how effectively the items used for the levels had been selected. Accordingly, the analysis was used for a while in conjunction with the spider diagrams, with items being discarded or shifted to neighbouring levels in the light of the number of error-types that this produced. However, it was gradually realised that the number of error-types was not just dependent on the strength of the relationship between the items used to form the levels, but on various other influences such as the number of levels formed and their differences in facility; moreover, the SPSS programme provided no adequate measures to compensate for these influences (nor, as far as is known, does anyone else). Thus similar difficulties were being met as with the inter-item measures (Chapter 9), except that the influences were even more complicated and the available measures behaved more like Y, Q and H_{ij} than like phi. It was therefore thought prudent to use scalogram analysis only to assess the practical value of the Algebra levels, ie only to determine the percentage of errors that occured when children's performance on the levels was described in terms of just the highest level passed. The errors obtained are discussed below, and the rest of the chapter looks at some of the difficulties involved in trying to assess their importance, with particular reference to the measures used in the SPSS programme.

The Percentage of Error-Types on the Algebra Levels

The table below shows the number of children passing each of the final Algebra levels, for the total 1976 Algebra sample (N=2923). The children and the levels are both ordered, in terms of the number of levels passed and the number of children passing; the error-types are shown in the body of the table, below the leading diagonal.

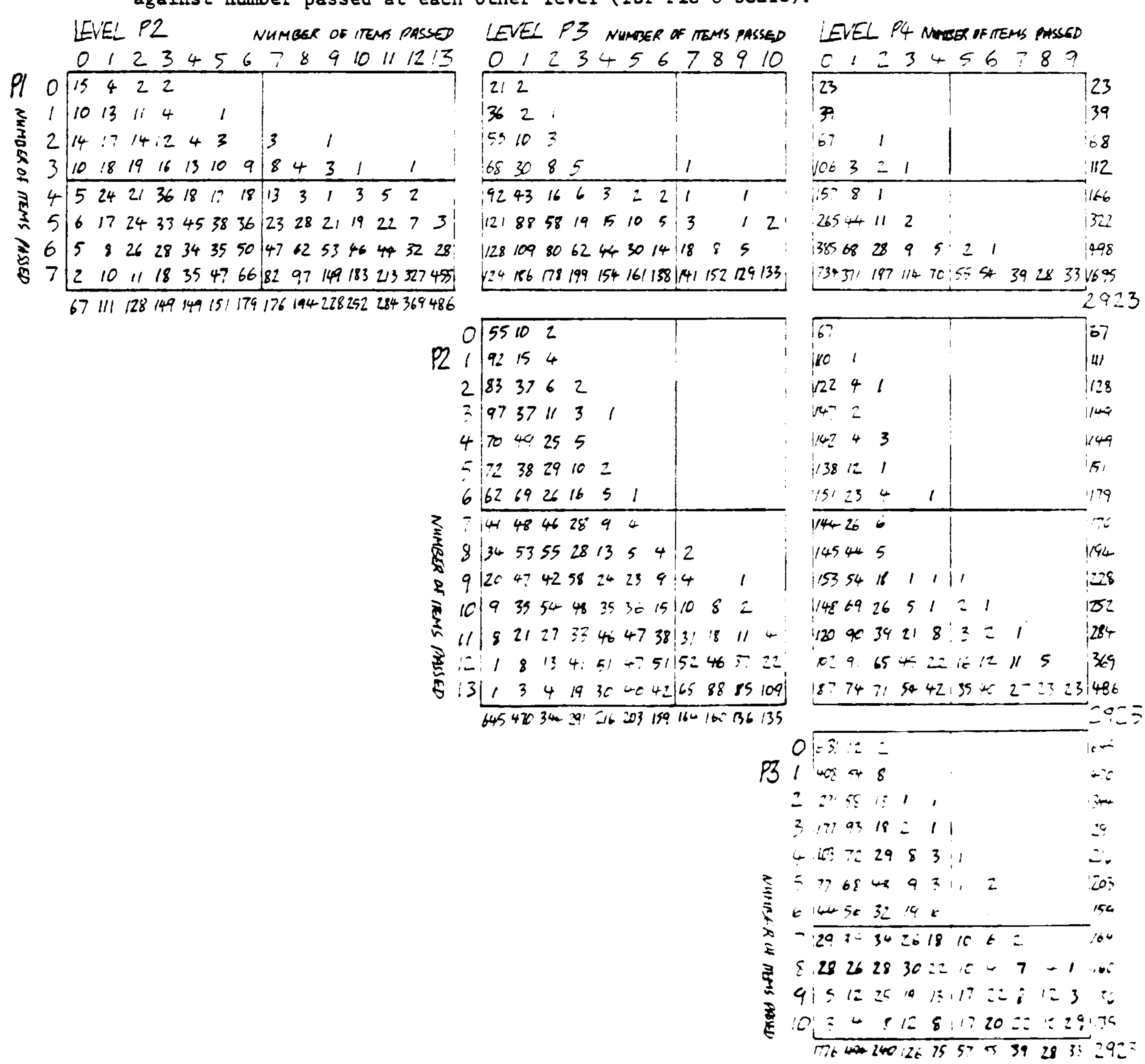
Fig 11.2 Scalogram for Algebra levels, on total 1976 sample.

	pass rate	Level 4 6/9 items	Level 3 5/8 items	Level 2 5/7 items	Level 1 4/6 items	total
number of levels passed	4	188	188	188	188	188
	3	1	717	718	718	718
	2	0	35	662	697	697
	1	0	0	0	1118	1118
	0	0	0	0	0	202
total		189	940	1568	2721	2923

As can be seen from the table, most of the errors occured between levels 3 and 2, with 35 children passing level 3 and level 1 but failing level 2. Altogether there were 36 error-types, which is just 1.23% of the total number of children.

On one of the earlier Algebra scales (PIG8) the number of error-types was even smaller (28, out of the 2923 children). However, here the pass-marks for the levels had not been set as close as possible to two-thirds of the items, but instead had been adjusted to make the marginal distributions match data obtained by the science wing. The following diagram was drawn for the scale, which shows the relationship between

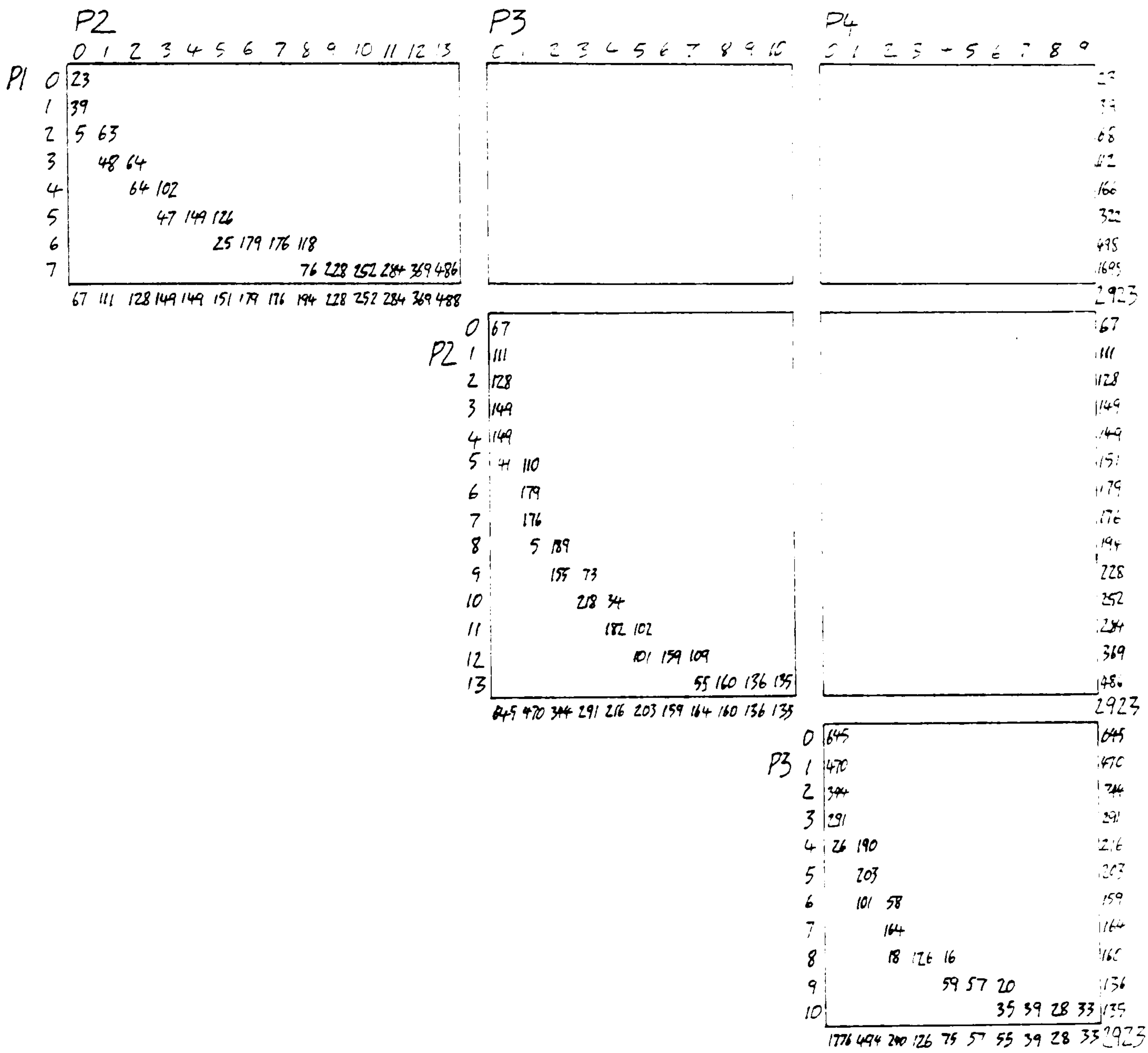
Fig 11.3 Frequency (N=2923) of number of items passed at one level, against number passed at each other level (for PIG 8 scale).



children's performance on the levels in considerable detail. Each table has been partitioned into four quadrants, according to the pass-mark used for the corresponding pair of levels, and the numbers in the upper-right quadrants show the number of error-types.

The above diagram is of interest one two counts. One, it shows that the error-types could be eliminated entirely, by judiciously shifting the pass-marks, even though the intrinsic relationship between the items would obviously stay the same. (It should be said that this was not done to any of the final levels of the CSMS mathematics tests, except, according to Hart (1981), in a very limited way in those cases "where the number of items in a group was not divisible by 3", when "the whole number above or below 2/3 was chosen, depending on which produced the better 'ordering'".) Second, regardless of the pass-marks, the diagram raises the question central to this chapter of how good the overall relationship between the levels that the diagram portrays can be said to be. In as much as the frequencies tend to be near the left-hand and lower edge of each table, the relationship seems, intuitively, to be quite strong. On the other hand, when the diagram is compared to the one below (which shows the ideal frequency-distribution for the given marginals, for P1 by P2, P2 by P3, and P3 by P4) it is clear that the relationship is not perfect. There are a number of measures that might be used for the pairs of levels separately (eg Pearson's r, though like phi this would suffer from the asymmetrical distribution of the marginals, or Kruskal's gamma which is a generalisation of Q); However, the point at issue is whether scalogram analysis provides a way of

Fig 11.4 Ideal frequency of number of items passed level x level, for given marginals (PIG 8 scale).



assessing the overall relationship between the levels, ie of assessing the consistency of the scale as a whole.

Measures of Scalability: Rep and Cos

Rather than simply determining the percentage of error-types, Guttman proposes a "coefficient of reproducibility" (rep) for measuring "the amount by which a scale deviates from the ideal scale pattern"; rep is evaluated

> "by counting up the number of responses which would have been predicted wrongly for each person on the basis of his scale score, dividing these errors by the total number of responses and subtracting the resulting fraction from 1" (Guttman,1950).

It should be noted that these "errors" are not the same as "error-types". Different authors unfortunately also count them in different ways. However, as far as the SPSS programme is concerned, the errors are determined by counting the number of ways in which an individual's response pattern differs from the ideal pattern for the same scale score. Some examples are shown below, for a scale consisting of 5 levels (or items).

Fig 11.5 Examples of error-count for SPSS programme GUTTMAN SCALE.

	LEVELS hard a	b	c	d	easy e	errors
observed response	1	0	1	1	1	2
ideal response	0	1	1	1	1	
observed response	1	0	1	0	1	2
ideal response	0	0	1	1	1	
observed response	1	0	0	0	0	2
ideal response	0	0	0	0	1	
observed response	1	1	0	0	0	4
ideal response	0	0	0	1	1	

As can be seen, the number of errors is always even, and for a scale consisting of only 4 or 5 levels (as with most of the CSMS mathematics tests) the number is rarely more than 2 per individual (ie per error-type). For the final Algebra levels shown earlier in this chapter, there were 36 error-types and 72 errors for the total 1976 sample. Here

$$\text{rep} = 1 - \frac{\text{number of errors}}{\text{total number of responses}}$$

$$= 1 - \frac{76}{4 \times 2923} = 0.993 .$$

The value of rep is 1 for a perfect scale, and 0 when the number of errors is equal to the total number of responses. However, in practice a value of 0 can not occur, because not all responses can be classed as errors. Infact the maximum number of possible errors is determined by the facilities of the levels (items), which means that this maximum varies from one scale to another. To take the simplest case, of a 2-level (or 2-item) scale, the number of error-types in this case is d (using the notation of Chapter 9) whose maximum value is equal to the minimum of p' and q (where p' and q are thought of as marginal frequencies, rather than percentages); the maximum number of errors is twice this. This situation is illustrated below, in the form of a contingency table and then as a scalogram, for levels (or items) whose facilities are 90% and 60% (cf Chapter 9, page 87).

Fig 11.6 Contingency table and scalogram for 90% by 60% facility items pair, when error-types are at a maximum.

			Item Q 0	Item Q 1
Item P	1	90	40	50
	0	10	0	10
			40	60

		Item P 0	Item P 1	Item Q 0	Item Q 1	Total
Scale Score	2	0	50	0	50	50
	1	40	10	10	40	50
	0	0	0	0	0	0
Total		40	60	10	90	100
Errors		0	10	10	0	
Total Errors						20

Here,

$$\text{rep} = 1 - \frac{10 + 10}{2 \times 100} = 0.90 .$$

Even when the marginals are all equal, rep has a lower limit of 0.50 for a 2-level scale.

Though Guttman claims that scalogram analysis "affords a rigorous test for the existence of single-meaning for an area"(ibid), he is aware of the inadequacy of rep, which he tries to compensate for by setting additional conditions for the acceptability of a scale. These concern the range of marginal distributions, the pattern of errors, the number of items and the number of response patterns; in particular, the scale should not just consist of extreme-facility levels (items) but should include levels of facility around 50%, and the number of levels in a scale should be at least 10 ("with perhaps a lesser number being satisfactory if the marginal frequencies of several items (levels) are in the range of 30 percent to 70 percent"(ibid)..). When all the conditions are met, Guttman suggests that a value of rep of 0.90 indicates an acceptable approximation to a perfect scale.

Understandably, many workers have not found this satisfactory, and have produced measures based not on a comparison of errors with the total number of responses but with some estimate of the maximum number of errors that can occur in practice, or the number of errors that would occur by chance (eg Menzel, 1953; Green,1956, and Loevinger,1948, whose index H is a generalisation of H_{ij}). Chilton (1966), using computer generated data, has estimated the chance distribution of rep, which allows the significance of an observed rep to be estimated, given a specific set of marginals and specified numbers of items and cases (as long as the data has not been manipulated to reduce the observed errors). However, for the large numbers in the CSMS samples, such a test of significance is irrelevant: as is noted by Green (1956,p86), "a significant rep does not necessarily indicate a homogeneous scale" (although the only additional advice that he can give is that the "intercorrelations of the items should be fairly high"..).

The dissatisfaction with Guttman's criteria arises from their subjective nature (Green calls the whole procedure "mystical") and the fact that "in practice these criteria are rarely met" (Chilton,1966) -which, with respect to the number of levels used, is true of the CSMS scales. However, it is still common to see scales described purely in terms of rep (eg Adi,1978, who regards a value of 0.96 as acceptable for a 3-level scale, and Versey,1978, who sees a value of 0.87 for a set of 25 tasks as an indication that the tasks are "not fully scalable").

In place of rep, the SPSS programme assesses scalability by comparing the errors with the "minimum marginal errors" (E_{mm}), (in a manner similar to, though cruder than, that used by Menzel,1953). E_{mm} is evaluated by summing the smaller of the marginals for each level. In the case of the final Algebra levels, which were answered correctly by 189, 940, 1568 and 2721 of the 2923 children,

$$E_{mm} = 189 + 940 + (2923 - 1568) + (2923 - 2721)$$
$$= 2686.$$

The "minimum marginal reproducibility" (rep_{mm}) is then given by

$$rep_{mm} = 1 - \frac{E_{mm}}{\text{total number of responses}}$$
$$= 1 - \frac{2686}{4 \times 2923}$$
$$= 0.770 .$$

Finally, scalability is assessed by comparing rep with rep_{mm}, or the observed errors (E) with E_{mm}, ie the "coefficient of scalability" (cos) is given by

$$cos = \frac{rep - rep_{mm}}{1 - rep_{mm}}$$
$$= 1 - \frac{E}{E_{mm}} = 1 - \frac{72}{2868} = 0.975 .$$

The rationale behind E_{mm} is the argument that for any single level, the maximum number of subjects that can be ordered wrongly (against some other criterion, ie against the order suggested by the scale as a whole) is p or p', whichever is the smaller (where p and p' are the number of subjects passing and failing the level respectively). Given $\cos = 1 - E/E_{mm}$, this in turn suggests that the value of cos ranges from 1 (for a perfect scale) down to 0 . However, as with rep, whose deceptive behaviour cos is meant to rectify, the value 0 is not necessarily attainable in practice. For example, for the two-level scale discussed on page 120, where the errors have the maximum possible value (E=10+10), the minimum marginal errors for the two levels are greater (E_{mm}=10+40), with the result that cos is substantially greater than 0 (cos = 1 - 20/50 = 0.60). Only when the facilities are equal can the value 0 be reached for a two-level scale.

In this respect, the behaviour of Loevinger's coefficient H is more satisfactory, in that its value is consistently 0, not when the errors are at their greatest but when they could have resulted from chance. To illustrate the behaviour of rep, cos and H, their values are shown below for a scale consisting of just two levels, of facility 80% and 60%, in the case of no errors, chance errors, and maximum possible errors. For clarity, contingency tables are used rather than scalograms.

Fig 11.7 Values of rep, cos and H for two-item scales, for items with 80% and 60% facility, for no, chance and maximum error-types.

	No Errors			Chance Errors			Maximum Errors		
	1	20	60	1	32	48	1	40	40
	0	20	0	0	8	12	0	0	20
		0	1		0	1		0	1
rep		1.00			0.88			0.80	
cos		1.00			0.60			0.33	
H (H)		1.00			0			-0.67	

Manipulating the Data: The Effects on Rep and Cos

In this section it is proposed to consider what happens to the values of rep and cos when a scale is altered, either by eliminating a level or by changing its pass-mark*. Specifically, it is proposed to examine scales derived for the CSMS Reflection and Rotation test. In common with the other CSMS scales, these scales contained relatively few levels and were made up of items which were already known to correlate reasonably well, with the result that the values of rep and cos were very close to 1, and any changes that occured in their values were small; nonetheless, it is of interest to consider the direction of the changes.

Changing the pass-mark of the levels

One of the Reflection and Rotation scales(PRRIG 1) was made up of 6 levels. The diagram below shows the errors obtained on each level (N=1026) when the pass-mark for each was set as close as possible to two-thirds of the items. The facilities of the levels for this specific scale (to be known as scale A) are plotted on the horizontal line.

Fig 11.8 Facility of levels and number of errors at each level for scale A (from Reflection and Rotation scale PRRIG 1).

Scale A

	Level 6		Level 5		Level 4		Level 3		Level 2		Level 1	
response	0	1	0	1	0	1	0	1	0	1	0	1
number of errors	0	15	4	6	15	9	10	33	33	5	6	0

One of the changes made to this scale was to raise the pass mark for level 1, ie to lower its facility (scale B). This

*Neither practice was used for any of the final CSMS levels, though according to Hart (ibid) error-types were not to exceed 7%, which meant that levels were sometimes amalgamated or items moved to a neighbouring level.

brought level 1 closer to level 2, from which it can be predicted that the errors would increase, which duly happened:

Fig 11.9 Facility of levels and number of errors at each level for scale B.

Scale B

	Level 6		Level 5		Level 4		Level 3		Level 2		Level 1	
response	0	1	0	1	0	1	0	1	0	1	0	1
number of errors	0	15	4	6	15	9	10	34	31	13	17	0

Since the total number of responses (6 x 1026) stays the same this increase in the number of errors means that rep should decrease. The situation for cos is more complicated: in theory, if cos fulfils the function it was designed for, it can be argued that its value should stay the same, since the new scale is made up of the same items as before; an examination of how cos is actually defined suggests that this may happen, because E_{mm} should increase as well as E, but the prediction is far from certain. In the event, rep behaved as predicted, but cos decreased, rather than staying the same:

Fig 11.10 Errors, rep and cos for scales A and B.

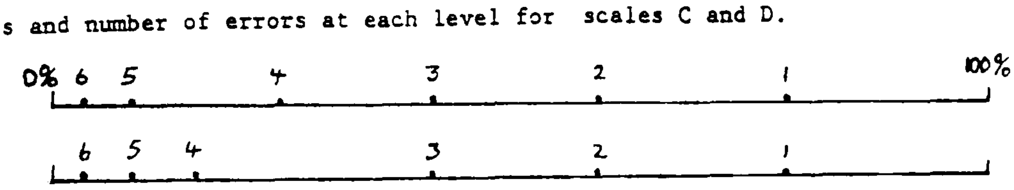
	errors	rep	cos
Scale A	136	0.978	0.892
Scale B	154	0.975	0.883 .

The diagram below shows the facilities of the levels and the corresponding errors, for two further scales, C and D, derived from PRRIG 1.

Fig 11.11 Facility of levels and number of errors at each level for scales C and D.

Scale C

Scale D



		Level 6		Level 5		Level 4		Level 3		Level 2		Level 1	
response		0	1	0	1	0	1	0	1	0	1	0	1
number of errors	C	0	11	5	4	8	32	25	31	34	11	17	0
	D	0	14	3	10	19	9	10	34	31	13	17	0

Compared to scale C, in D the facility of Level 4 was decreased, moving it away from level 3 and closer to level 2. At first sight this suggests that the total number of errors might stay the same, with the errors decreasing between levels 3 and 4 and increasing between levels 4 and 5. However, level 3 original produced more errors than level 5 (perhaps because it is a less homogeneous level, but also because its marginals are more nearly equal) which means that there is likely to be a greater decrease in errors between levels 3 and 4, than the increase between levels 4 and 5, which is how it turned out. This means rep should increase, which it did, and again cos should tend to stay the same (because the changes in E and E_{mm} are once more in the same direction), but this time it increased, as is shown in the table below.

Fig 11.12 Errors, rep and cos for scales C and D.

	errors	rep	cos
Scale C	178	0.971	0.875
Scale D	160	0.974	0.880

Another transformation that was applied, which is in many ways similar to the one above, was to shift level 5 closer to level 6. However, this time level 5 was already sufficiently far away from level 4 and close to level 6 for the decrease in errors resulting from moving away from level 4 exactly to balance the increase from moving closer to 6. This means that rep stayed the same but cos decreased; however, it also means that if level 5 were moved closer still to level 6, a further decrease in cos would be almost inevitable.

Reducing the number of levels

On the assumption that the levels forming a scale differ in homogeneity, it would seem reasonable to suppose that the effect on cos of removing one of the levels would depend on which level is chosen. It therefore came as a surprise when, using a 5-level scale GRROsna, the value of cos increased whichever level was removed. Though Guttman warns about having too few levels when it comes to interpreting rep, it had been assumed that for cos this would not matter.

Fig 11.13 Cos for scale GRRCsna when one level removed.

Scale GRROsna Single Level Removed	None	Level 5	Level 4	Level 3	Level 2	Level 1
Resulting Value of Cos	0.891	0.899	0.906	0.940	0.953	0.904

On further investigation it became clear that the effect on cos was not the result of using a freak scale, but was likely to happen with many scales, and for some scales would be inevitable. This can be explained best by writing the quotient E/E_{mm} in the formula for cos ($cos=1-E/E_{mm}$) in the following way:

$$\frac{E}{E_{mm}} = \frac{E_5 + E_4 + E_3 + E_2 + E_1}{E_{mm5} + E_{mm4} + E_{mm3} + E_{mm2} + E_{mm1}},$$

where E_i and E_{mmi} are the errors and minimum marginal errors for level i (for, in this case, i=1 to 5). At first sight, it would seem that the effect on E/E_{mm} of removing level 5, say, would be to remove (subtract) E_5 and E_{mm5} from the numerator and denominator respectively. If this were so, and if also the quotients E_i/E_{mmi} (for i=1 to 5) were all equal (and equal to E/E_{mm}) there would be no effect on the value of cos. However, because the errors in one level cause errors in other levels, the effect of removing level 3, say, would be to remove a greater number of errors than E_3. Therefore, if the quotients E_i/E_{mmi} are all equal, E/E_{mm} would decrease and cos ($1 - E/E_{mm}$) would consequently increase.

In practice, the quotients E_i/E_{mmi}, though not equal, are likely to be similar, and cos is therefore likely to increase on the removal of one of the levels. It is only when one of the quotients is very much smaller than the others (and in particular when E_i for one of the levels is zero) that cos is likely to (will) decrease. For the scale GRROsna, E/E_{mm} was equal to 0.109, and the quotients for levels 5 down to 1 equalled 0.125, 0.107, 0.104, 0.116 and 0.102 respectively. On the other hand, the quotients formed from the number of errors actually removed and the minimal marginal errors removed on the removal of any single level were 0.250. 0.214. 0.203, 0.233 and 0.192 respectively. These numbers are about twice the size of E/E_{mm} and E_i/E_{mmi}, which stems from the fact that when a level is removed about twice as many errors are removed as are under the the level itself. This is shown in the two tables below.

Fig 11.14 Errors and errors removed for scale GRROsna when one level removed.

	Level 5		Level 4		Level 3		Level 2		Level 1		
response	0	1	0	1	0	1	0	1	0	1	Total
number of errors	0	8	7	10	11	32	31	16	17	0	132
level 5 removed	-	-	0	10	10	32	31	16	17	0	116
level 4 removed	0	1	-	-	1	32	31	16	17	0	98
level 3 removed	0	8	8	0	-	-	0	16	16	0	48
level 2 removed	0	8	7	10	11	1	-	-	1	0	38
level 1 removed	0	8	7	10	11	32	32	0	-	-	100

	Level 5	Level 4	Level 3	Level 2	Level 1
total errors for each level	8	17	43	47	17
errors removed when level removed	16	34	84	94	32

The numbers in the tables can be derived (tediously) from the networks below, which show how many children passed each level and which easier levels they passed (by moving up the networks). Also shown is the scalogram for all 5 levels of scale GRROsna, for N=1026 (from Küchemann,1978a).

Fig 11.15 Networks showing number of children at each level for scale GRROsna when one level removed.

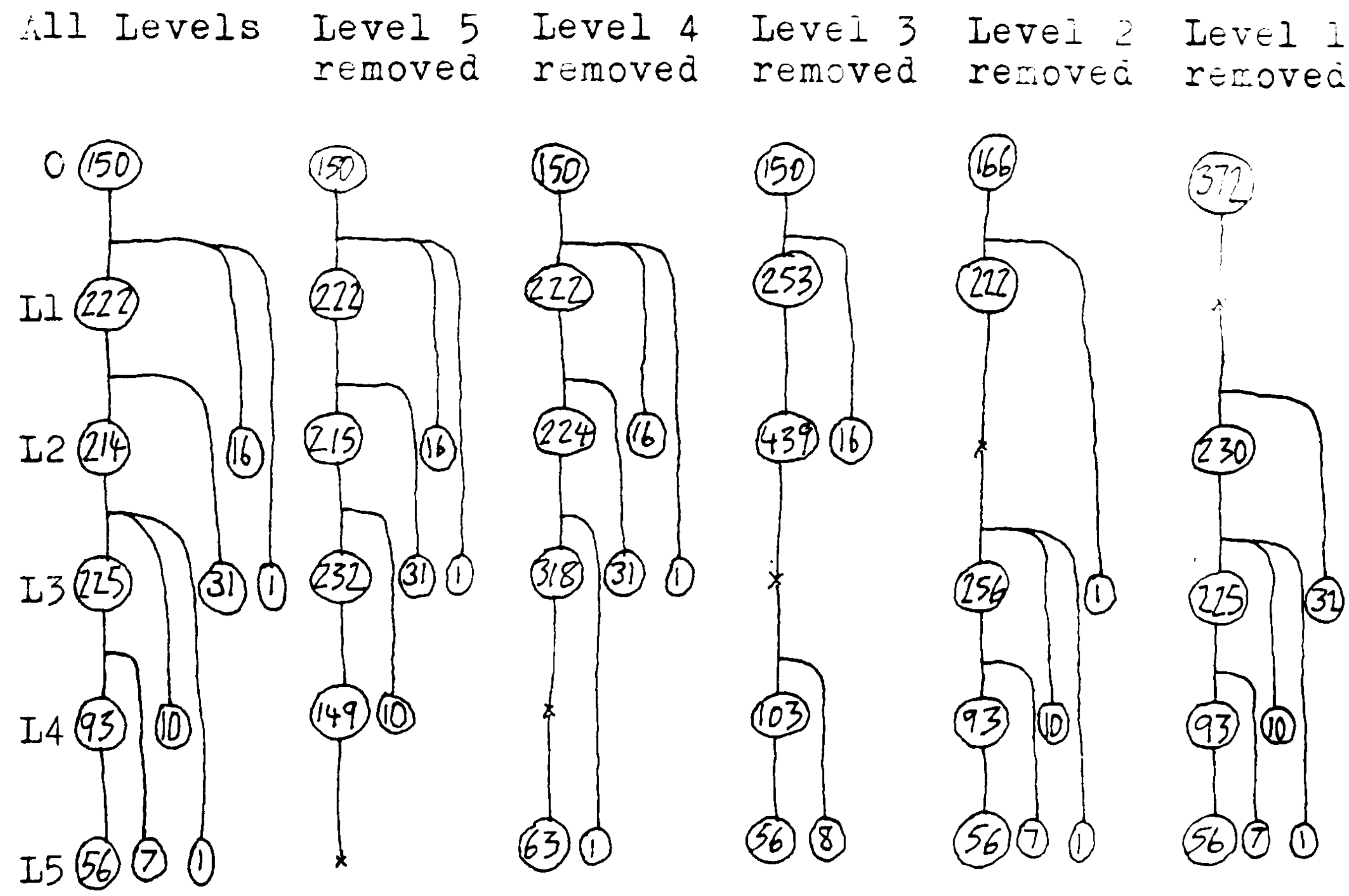


Fig 11.16 Scalogram for scale GRROsna.

Scalogram for scale GRROsna (correct responses only)

pass rate		Level 5 9/13 items	Level 4 5/7 items	Level 3 5/8 items	Level 2 4/6 items	Level 1 7/11 items	total
SCALE SCORE	5	56	56	56	56	56	56
	4	7	93	100	100	100	100
	3	1	10	225	236	236	236
	2	0	0	31	214	245	245
	1	0	0	1	16	222	239
	0	0	0	0	0	0	150
tot		64	159	413	622	895	1026
E_i		0+ 8	7+10	11+32	31+16	17+ 0	132
E_{mmi}		64	159	413	404	167	1207

$$\text{rep} = 1 - \frac{132}{5 \times 1026} = 0.974$$

$$\cos = 1 - \frac{132}{1207} = 0.891$$

Chapter 12 THE ALGEBRA LEVELS AND PIAGETIAN THEORY

Seen in a kindly light, the problem of the validity of Piagetian theory is a fascinating one, and some of the evidence and arguments about it are considered later in this chapter. However, in relation to the aims of the mathematics wing of CSMS the problem is not very real: the question of whether it is worthwhile trying to describe children's understanding of mathematics, and in particular of algebra, in Piagetian terms is as open now as before the CSMS mathematics work started and has nothing to do with the validity of the theory but with its practical utility. Shayer (1978c) has shown that assessing children's Piagetian level of cognitive development can provide a quite accurate predictor of their success on science courses analysed in the same terms; however, in mathematics this work still needs to be done.

A subsidiary question is whether children's understanding of mathematics (or the cognitive demand of mathematics tasks) can be described in Piagetian terms, which is in one sense answered by someone deciding to do so, although the quality of the end-product and the likelihood of it being accepted will depend on what guidelines are available. As far as secondary school mathematics is concerned, Piaget's own work is not very helpful in this respect, beyond providing very general descriptors (such as the ability to use second-order operations and to think in terms of hypotheses and possibilities at the formal operational stage, instead of being bound be concrete reality). The work of some "neo-Piagetians" is of interest and this will also be considered later in this chapter. However, first it is proposed to discuss some empirical comparisons between the Algebra test and Piagetian tasks.

Empirical Comparisons

Children from two schools (11 and 14) included in the 1976 Algebra survey, also took the Pendulum task (Küchemann,1979), which is one of the Piagetian class tasks developed by the science wing of CSMS, on the basis of the descriptions given in Inhelder and Piaget (1958,Chapter 4). Children's performance on the two tests is shown by the table below (N=248).

Fig 12.1 Algebra levels by Pendulum levels (N=248).

Algebra	42	66	84	30	26	248
Level 4		1	4	3	10	18
Level 3	7	16	33	15	14	85
Level 2	13	17	24	8	2	64
Level 1	18	30	20	4		72
0	4	2	3			9
	<2B	2B	2B3A	3A	3B	

Pendulum

The product-moment correlation between the two tests was 0.49. On the other hand, Shayer independently classified the Algebra test into the same number of levels as the Pendulum task, using Piagetian descriptors and swags, and obtained a correlation of 0.58 on a subsample of the children (N=127):

Fig 12.2 Shayer's Algebra levels by Pendulum levels (N=127).

Algebra	15	35	42	20	15	127
3B			2	2	5	9
3A	1	2	8	8	8	27
2B3A	2	12	13	6	2	35
2B	8	9	17	2		36
<2B	4	12	2	2		20
	<2B	2B	2B3A	3A	3B	

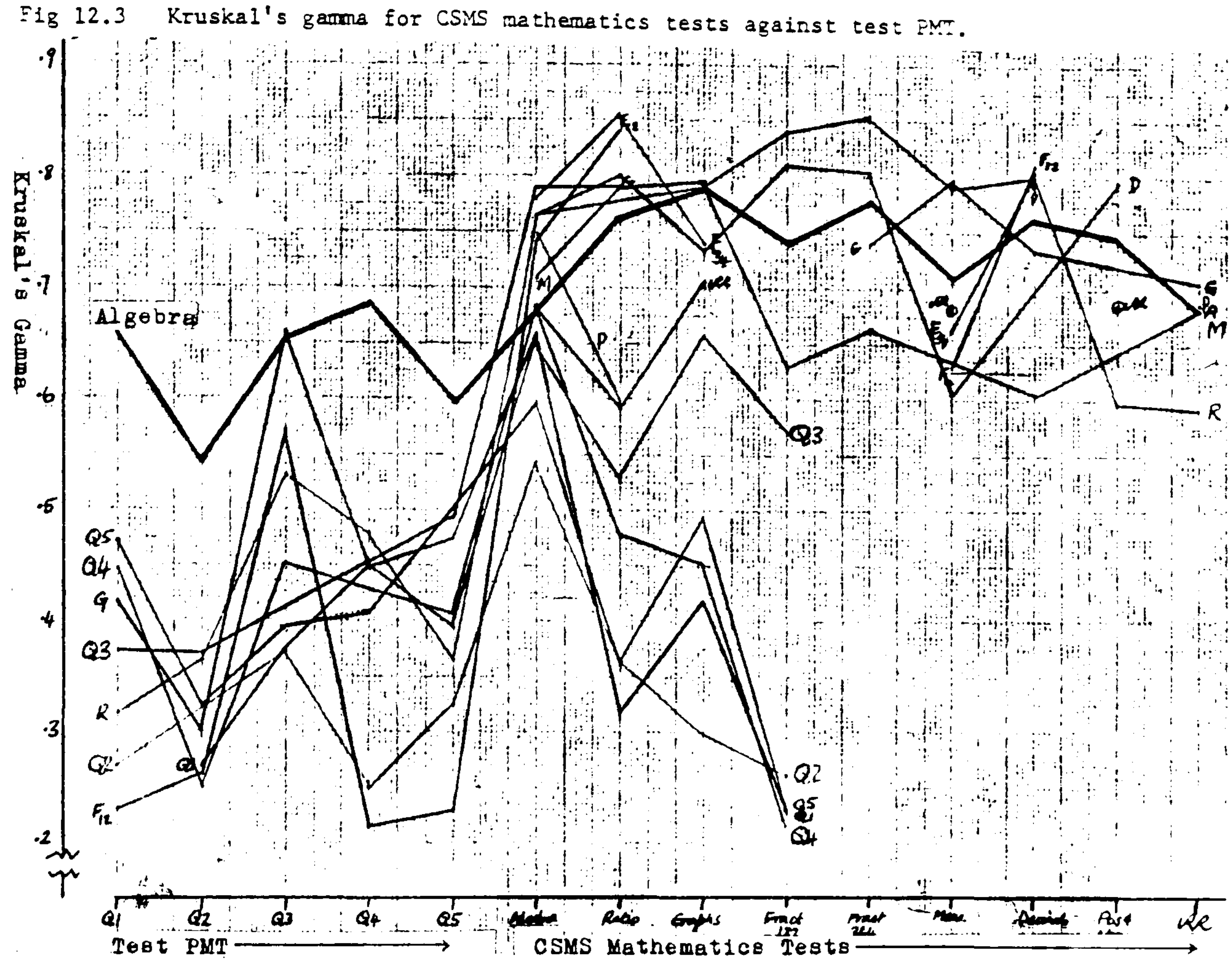
Pendulum

The value 0.58 is the same as the mean between Pendulum and four other Piagetian class tasks (Shayer,1979), but both the obtained values are lower than the correlations between the Algebra test and the other CSMS mathematics tests (which had a mean of 0.65, with eight other tests). Also the correlations (using Kruskal's gamma) between the Algebra test and the five questions of a Piagetian test (test PMT) devised by the mathematics wing of CSMS (see Hart,1980) are lower than between Algebra and the other mathematics tests, as can be seen from the graph* below -although this may partly be due to the questions being very much shorter than the tests.

However these correlations are judged, as far as those between Algebra and Pendulum are concerned it can be argued that they are sufficiently high to give credence to the Piagetian substages assigned to the items by Shayer in his analysis;

* [illegible] by M McCartney

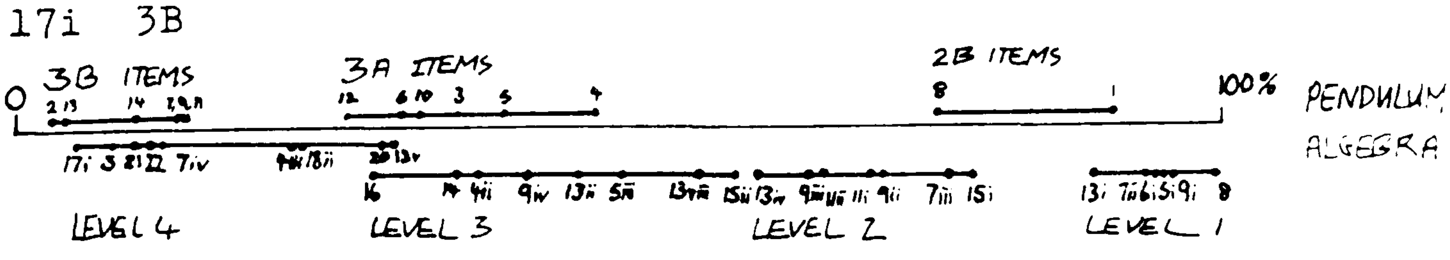
Fig 12.3 Kruskal's gamma for CSMS mathematics tests against test PMT.



also, though it is not possible to make a direct match between the final Algebra levels and the levels of the Pendulum task, it can be argued that the correlations are high enough to match the individual items by means of their facilities. The table below shows the substages resulting from Shayer's analysis, for the 30 items used for the final Algebra levels, whilst the graph

Fig 12.4 Piagetian substages of selected Algebra items (Shayer), and facility of Algebra and Pendulum items and Pendulum substages.

Level 4	Level 3	Level 2	Level 1
18ii 3B	15ii 2B+	15i 2B	8 2A
13v 3A	13viii 3A	9ii 2B	5i 2A
20 3A	13ii 3A	7iii 2B	9i 2A
4iii 3A	5iii 3A	9iii 3A	6i 2A
7iv 3B	14 3B	11ii 2B	7ii 2A
21 3B	9iv 3A	11i 2B	13i 2A
22 3B	4ii 3A	i3iv 2B	
3 3B	16 ?		
17i 3B			



shows the facilities of these items and those of the Pendulum task on the same subsample of 248 children. Also shown on the graph are the substages of the Pendulum items (taken from Shayer,1978a,p192). The swags used by Shayer in the Algebra analysis are shown in Appendix 12.1, and the Pendulum task is shown in Appendix 12.2.

There is quite close agreement between the table and the graph, and together they suggest that the level 4 items (at least the harder ones) can be assigned to the substage of late formal operations (3B), the level 3 items to early formal operations (3A), with some of the easier ones perhaps at the transition to early formal operations, whilst the level 2 and 1 items are at the concrete operational stage.

The ways in which the letters are interpreted at the different levels (see Chapter 5) lend support to this classification. For example, having to evaluate letters, or ignoring them or using them as objects (levels 1 and 2) seems very much to indicate a reliance on directly intuitable reality (concrete operations), whereas operating on unknowns, operating on operations, and forming relations between relations (levels 3 and 4) are each regarded as characteristics of formal thought, by, for example, Collis (1974), Halford (1978) and Inhelder and Piaget (1958,p254) respectively.

Validity of Piagetian Theory

Many of the arguments about the validity of Piagetian theory can conveniently be considered under the four headings "task variations", "training studies", "meta-theory" and "unity of the stage-construct".

With respect to the first of these, a growing number of studies have shown that by changing some of the features of classical Piagetian tasks, children are able to demonstrate some grasp of certain fundamental concepts at a substantially younger age than stated by Piaget: eg invariance of number (Gelman,1972), transitivity (Bryant and Trabasso,1971; but see De Boysson-Bardies and O'Regan,1973 and Halford and Galloway,1977 for an alternative explanation for Bryant and Trabasso's findings), seriation, class inclusion, coordination of viewpoints, etc, etc (Donaldson, 1978). These studies cause problems for Piagetian theory in as much as they blur the distinction between one operational stage and the next, and as such they underline the importance of defining the precise nature of any task used to assess children's cognitive development and of defining what is meant by "success". However, it is difficult to see how their findings contradict Piagetian theory, as they are often claimed to do (eg Entwistle, 1979). To quote Pascual-Leone (1976,p271):

> "The fact that Piaget's structuralist theorising has usually neglected this issue and that he has written as if competence markers were context free, is no excuse for critics to do the same".

With respect to training studies, a series of "naturalistic" studies undertaken by Inhelder et al (1974), which employed cognitive conflict, showed that children could make substantial advances in their understanding of a given task, but only if they had already reached a certain minimum level of cognitive development. In contrast, Brainerd (1978) claims that there is extensive evidence that studies employing more "direct" methods (such as "simple correction", "rule learning", "observational learning" and "conformity training") are far more effective, regardless of the child's developmental level.

If such methods can be shown to have genuine and long-term effects on children's cognitive structures, this would have important implications for Piagetian theory, and for teaching, since there would then be little point in trying to assess a child's cognitive level with a view to matching the teaching to it. However, as Pascual-Leone (1976) has argued for one such study, the success of the studies cited by Brainerd seems primarily to be due to children overlearning a specific rule or strategy (such as "ignore length because it's unreliable") rather than their developing more complex cognitive structures (as indicated, say, by the recognition that an increase in length of a row of beads is compensated by a decrease in density). Thus it is of interest to note the explanations that children gave in these studies, though unfortunately this information is not always forthcoming: Zimmermann and Rosenthal (1974), who trained children to conserve equalities and inequalities, give no indication of whether children's explanations involved identity, reversibility or compensation; neither does Murray (1972), though he goes as far as to say that "children's reasons indicate that in some sense reversibility and indentity rules were acquired"(ibid,p5), rather than compensation, and that, for example, the reversibility explanation for conservation judgement is "strictly speaking incorrect". Gelman(1969,p183) states that in her study such explanations "occured frequently". Sheppard's reporting is more explicit: after training "75% of the explanations emphasise(d) identity" (Sheppard,1974,p727); also, in contrast to the other studies cited by Brainerd, he found that the initial level of operativity was of significance in respect of whether children benefited from the training.

Shayer (1979) uses the term "meta-theory" for Piaget's attempt to explain the developmental stages in terms of logico mathematical structures, and in particular for Piaget's use of propositional logic to describe formal operational thought (in Inhelder and Piaget, 1958). Parsons (1960) has shown that this use of the propositional calculus is unorthodox and obscure; however, this does not necessarily invalidate the meta-theory and, as Shayer (ibid) points out, it certainly does not follow that the "lower tiers" of Piaget's theory (the behavioural descriptions, and their classification into stages) is unsound, as Parsons seems to imply. Ennis's argument (1975) that the lower tiers are unsound because they cannot be deduced from meta-theory is equally untenable.

Ennis also argues that Piaget's claim that only children at the formal operational stage can cope with propositional logic is refuted by the fact that much younger children can sometimes cope with items like the following:

If this is room 9, then it is the fourth grade.
This is room 9.
Is it the fourth grade?

However, as Knifong (1974) points out, this logical form (modus ponens) allows a definite inference which can be made using "transductive reasoning". It is with the inverse and converse forms, where no definite inference can be made, that children and even adults (eg O'Brien,1973) have difficulties, and it may well be that these do require formal thought. On the other hand there is abundant evidence (eg Johnson-Laird and Wason,1970) that the logical "selection task" devised by Wason, in its original "abstract/arbitrary" form, is substantially more difficult than the formal operational tasks devised by Inhelder and Piaget. This indicates that the grasp of formal logic displayed by children at the formal operational stage is not as complete or as flexible as Piaget's meta-theory suggests. At the same time, Wason's task is very different from Piaget's formal tasks: Wason's task requires an understanding of the "falsification principle" -the conditions necessary to test the truth or falsehood of a given rule- whereas the primary aim in the formal tasks is to find a rule, whereby falsification only becomes important in the sense of controlling for extraneous variables: for example in the Flexible Rods task, if a child has formed the hypothesis "If brass, then more flexible (than iron, say)" $(B \rightarrow F)$, the child needs to be aware that $B.\bar{F}$ (A less flexible brass rod) may exist, as a result of choosing a thicker rod perhaps, and that he therefore has to compare rods of the same thickness; but, his main aim is to confirm the hypothesis, ie to seek out B.F. Moreover, the formal tasks are in a realistic context, and in this respect it is very interesting to note that when a more plausible rule was used in the selection task (eg "Every time I go to Manchester I travel by train", Wason and Shapiro,1971, or "If a letter is sealed, then it has a 50 lire stamp on it", Johnson-Laird et al,1972, rather than the arbitrary "If a card has a vowel on one side, then it has an even number on the other side", Wason,1966;) the task was very much easier.

The work of Wason and Johnson-Laird does not mean that the meta-theory should be rejected out of hand: at the least it can be regarded as a useful heuristic device. Nor does their work refute the lower tiers of Piaget's theory. Rather, the importance of their work, in relation to Piagetian theory, is to have removed some of the misconceptions about formal thought that Piaget, in his attempt to construct a meta-theory, himself introduced, and that has led to statements such as "formal operations ... demand the manipulation of abstract concepts" (Entwistle,1979,p124).

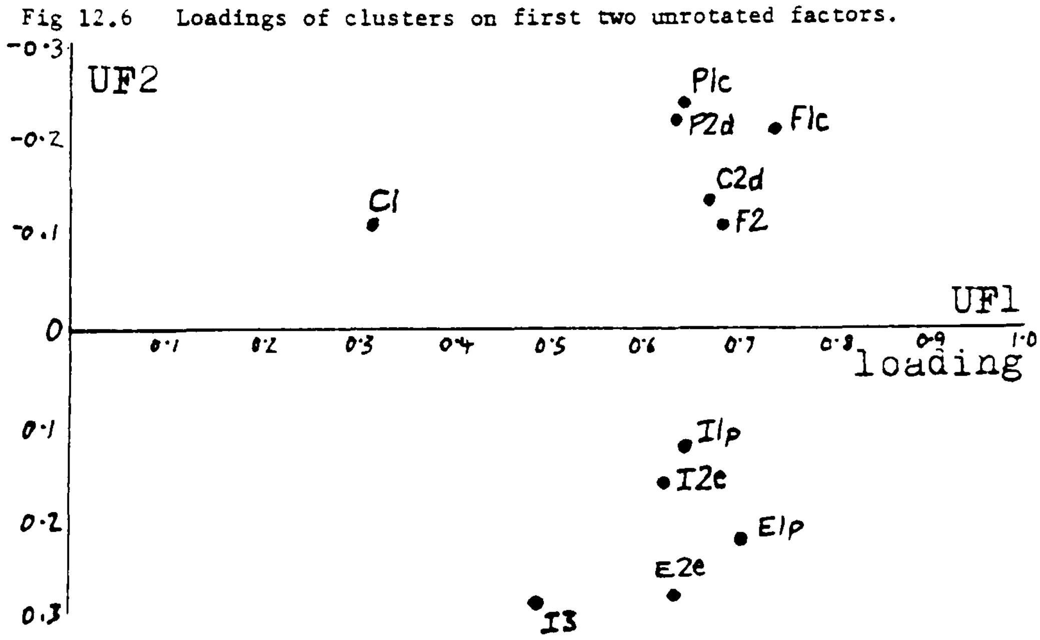
As far as the unity of Piaget's stage-construct is concerned, one of the most widely quoted criticisms is the review by Brown and Desforges (1977) who argue that the available correlational data show the notion of stage to be untenable. However, their choice of data was highly selective, as is shown by Shayer (1979), and also by their unwillingness (Desforges and Brown,1979) to defend the arguments they had used. Shayer demonstrates that when the correlations between formal tasks that are reported in the literature (including those cited by Brown and Desforges, which had values of about 0.3 to 0.4) are adjusted to match the distribution of the sample tested by Shayer himself (which consisted of a "representative sample of adolescents taken from one year group") the values are of the order of 0.6 , as are the values that Shayer found between five of the CSMS Piagetian class tasks. This ofcourse still leaves the problem of deciding whether these values are sufficiently high to regard the construct of a formal operational stage as valid, which Shayer attempted to solve by factor analysing the CSMS tasks, separately and then together. The separate analyses split each test into two or three clusters of items, which, by inspecting the items, Shayer described in terms of the Piagetian schemas shown below (from Shayer,1978a,p188):

Fig 12.5 Description of item clusters for CSMS Piagetian class tasks.

Task	Number of items in cluster	Description of cluster	
Pendulum	5 items	"control of variables"	P1c
	5 items	"deduction of effects"	P2d
Equilibrium in the Balance	3 items	"proportionality"	E1p
	9 items	"equilibrium of system"	E2e
Inclined Plane	5 items	"proportionality"	I1p
	6 items	"equilibrium of system"	I2e
	3 items	"work interpretation"	I3
Chemical Combinations	4 items	"combinations"	C1
	9 items	"strategy and deduction"	C2d
Flexible Rods	7 items	"control of variables"	F1c
	9 items	"compensation"	F2

It can be seen that the same description sometimes occurred for clusters from different tests; thus "control of variables"(c), "deduction"(d), "proportionality"(p) and "equilibrium"(e) each appeared twice. The clusters were then treated as individual variables and put into one factor analysis (N=428). Initially, only two of the resulting factors (components) had eigenvalues greater than 1, and these accounted for 44% and 9% of the total variance, with the next two components accounting for 8% and 7% respectively. Using the "Cattell Scree test", Shayer suggests that only the first component is significant, although the computer programme (PA2; SPSS,1975) selected the first two for further analysis, the results of which are shown in the graph below (where UF1 and UF2 are the two unrotated factors).

Fig 12.6 Loadings of clusters on first two unrotated factors.



As can be seen from the graph, there is a tendency for variables from the same task to cluster together, in much the same way as items from the same question did on factor analysis of the Algebra test (Chapter 10). Also the variables representing the same schemas, but from different tests, (ie Plc and Flc, P2d and C2d, Elp and Ilp, E2e and I2e) tended to cluster, and it is interesting to speculate whether, if there had been more variables than just two for any one of these schemas, the relationship between the variables would have been strong enough to produce and "independent" factor (ie an initial component with eigenvalue greater than 1, which is unlikely to happen for just two variables). However, as it stands, Shayer is clearly right in saying the analysis produced only one significant factor (the second factor only exists to accomodate the partially conflicting tendencies mentioned above); put another way, the evidence from the factor analysis supports (does not refute) the claim that Piaget's formal operational stage, made up of the schemas listed above, is a unitary construct. The only exception is the "combinations" schema, whose loading on UF1 in particular is much less than that of any of the other schemas.

In summary, it has been argued that evidence against the validity of Piagetian theory in less compelling than some critics have suggested. However, the theory does lack an adequate meta-theory that could be used deductively to assess the cognitive demand of untested areas of the curriculum. Also, though children's performances on different tasks are more consistent than Brown and Desforges suggest, data from correlational studies indicate that classifying children into cognitive developmental stages provides only a first approximation of their abilities on different tasks.

As far as the utility of the theory is concerned, Shayer (1978c) has shown that this degree of approximation is sufficient to make it worthwhile analysing science curricula in Piagetian terms, in order to devise curricula that are more suited to the individual child. However, it has yet to be demonstrated that the attempt to match children's cognitive levels to the cognitive demand of mathematics is worth undertaking, and before it could be done a detailed taxonomy would need to be devised for analysing mathematical tasks. A start might be made by forming empirical comparisons, of the kind mentioned in this chapter; on the other hand, the complexities of mathematical tasks are such that it might be considered more fruitful to concentrate on identifying individual dimensions that affect task difficulty (see the next chapter), rather than trying to assess their interaction in the hope of assigning tasks to an overall level.

Neo-Piagetian Theories

Attempts have been made to reinterpret Piaget's theory in terms of the development of short-term memory capacity, by determining the number of elements (or concepts or schemes) that have to be considered simultaneously to solve tasks at the different stages. McLaughlin (1963) proposed that preoperational, concrete operational and formal tasks involve a coordination of 2, 4 and 8 (2^1, 2^2 and 2^3) concepts respectively, and he suggests that a child would need an equivalent digit-span to be at a given stage. Halford (1978) also argues that a span of 2 is required for preoperational tasks and 4 for concrete operational tasks, but he suggests that formal tasks require a span of 6 rather than 8. In the theory developed by Pascual-Leone and his coworkers (see for example Case, 1974) the emphasis is on processing rather than storage capacity, and thus, rather than using a straightforward span test, the child is tested on recall of information that first has to be operated on in some way: for example, the child might be asked to recall the number of elements in several sets which he has first had to count. According to Case (1978a), children below the age of 5 years can only remember the number of elements in one set, whereas by, say, age 7 or 8 they can remember the number of elements in as many as 3 sets. The table below (from Case,1974) shows the relationship, according to Pascual-Leone's theory, between age, developmental level and processing capacity ("M-power"). The constant e refers to the mental effort required to attend to the specific instructions of the task (when these are familiar), whilst the numeral represents the maximum number of schemes that have to be considered simultaneously to solve a task at a given Piagetian substage, whose value, for an individual child, can be determined by the kind of test referred to above.

Fig 12.7 Relationship between age, developmental level and processing capacity, according to Pascual-Leone's theory.

Age	Developmental substage	M-power
3-4	Early preoperations	e + 1
5-6	Late preoperations	e + 2
7-8	Early concrete operations	e + 3
9-10	Middle concrete operations	e + 4
11-12	Late concrete-early formal operations	e + 5
13-14	Middle formal operations	e + 6
15-16	Late formal operations	e + 7

There is a certain amount of empirical evidence to support Halford's and especially Pascual-Leone's theory (eg Halford, 1978; Case,1974,p584,footnote 4). However, at present, their practical value would seem to be limited: the analysis required to determine the schemes relevant to a particular task can be extremely complicated (eg Pascual-Leone and Smith,1969) and ambiguous (eg Lawson,1976) and the resulting estimate of task difficulty can turn out to be very broad (eg Halford,1978). Also, the fact that it is possible to find some kind of correspondence between, say, digit span and success on tasks that have been assigned to a particular cognitive level (eg Collis,1975a,chapter 7) is not really very surprising, and there is a danger that its importance is being over-estimated. Moreover, there is a danger that too literal a link is being made between span and cognitive level: though performance on span tests certainly improves with age, there is abundant evidence to suggest that the corresponding growth in short-term memory capacity, or in processing capacity, can not be seen in simple numerical terms -for example as an increase in the number of discrete "memory slots": as well as depending on age, performance on span tests has been shown to depend on the familiarity of the elements being recalled (eg Crannell and Parrish,1957), and possibly also on the level at which the elements are being processed (eg Craik and Lockhart,1972), the child's ordering ability (Huttenlocher and Burke,1976), and the strategies available to him, such as rehearsal (eg Kingsley and Hagen,1969). Case (1978b) seems to have modified his views in accordance with some of these findings by suggesting that

> "the gradual increase in working memory does not stem from a structural increase in the attentional capacity of the organism, but rather from an increase in the automaticity of the basic operations it is capable of executing. As these operations become more automatic, their execution requires a smaller proportion of the attentional capacity. The result is that more capacity is available for 'storage' or 'working'".

Halford ignores the element dimension (ie the degree of familiarity) and instead "attempts to subsume the various characteristics of concrete and formal reasoning under the single factor of operational complexity" (Halford,1978,p298). However, this dimension is precisely what the CSMS Algebra research is all about, and it therefore comes as no surprise that Halford cites items classified at the same operational level that are vastly different in facility. For example, the two item-types shown below are both classified by Halford as formal ("two operations in each expression"), but Halford reports that the proportion of correct responses for a set of items of the first sort was 0.87 (for 36 11 to 12 year olds), compared to only 0.42 for a set of the second sort.

Fig 12.8 Item types classified by Halford as formal.

First item-type:	$\frac{6 \times 4}{2}$, $\frac{6 \times 8}{4}$	(Are these expressions the same? different? or is it impossible to tell?)
Second item-type:	$\frac{a \times b}{c}$, $\frac{a \times 2b}{2c}$	

Halford's only comment is:

> "Although theoretically the same process in required .., it is possible that most children find it harder to select the appropriate strategy when presented with unknowns" (ibid,p307).

Collis does take note of the element dimension. For example, in one study he contrasts children who try to solve the item "Find y if y=b and y+2b=90" by giving numerical values to b, and who "eventually get lost in a series of trials", with children at a higher level of cognitive development, who argue in the following, much more efficient manner: "b is a number; 2b is twice that number and thus twice y; y and 2y make 3y and thus y=30" (Collis,1975b,p46). There is an interesting paradox in this example, in that the second group of children, who can be assumed to have a greater "storage", "working" or "processing" capacity, choose a strategy which is less complex than the trial and error method of the first group of children. As far as their willingness to operate on the letters is concerned, it may be that the second group are more "experienced" in this use of letters; however, it can also be argued that the ability to use letters as numerical entities in their own right, or rather, learning to use letters in this way, requires a greater processing capacity than working just with numbers: for example, children may initially have to invoke known numbers and the way they behave in support of their efforts to work with letters per se. This raises an interesting possibility, in algebra and other areas of mathematics, that an increase in processing capacity may give children access to strategies that are not only more advanced but that, once understood, require less processing capacity than the strategies they displace; put another way, there may be a "threshold effect" on the learning of new strategies, which an increase in processing capacity overcomes.

Although these neo-Piagetian theories are difficult to apply in practice, they do make the important point that the complexity of a task (and the degree of familiarity of its elements) has a crucial effect on task difficulty. Also, they suggest there may be a limit to the kind of task that a child can be taught to solve or understand at a given time.

Chapter 13 ALGEBRA AND THE OTHER CSMS MATHEMATICS TESTS

This chapter compares the facilities of selected items within and across the CSMS mathematics tests*, in order to examine some of the dimensions that affect the difficulty of mathematics tasks generally. Only items used to construct the levels will be compared, so as to avoid items that behaved in an inconsistent way. Generally the inter-item correlation within levels was about 0.4 (using phi) and, as the table below shows, the correlation between tests was of the order of 0.6 (using Pearson's r). These values suggest that the relationship between items, within and across tests, is strong enough for a comparison of facilities to be undertaken. To simplify the discussion, the facilities will be those from the 3rd year samples.

Fig 13.1 Correlation (Pearson's r) between CSMS mathematics tests (Vectors and Matrices tests not included).

Ratio	.66						
Graphs	.60	.60					
Fractions 3&4	.71	.69	.65				
Measurement	.61	.65	.71	.66			
Decimals	.73	.70	.65	.78	.60		
Integers	.68	.48	-	-	-	.77	
Reflection & Rotation	.61	.54	.60	-	.63	-	.61
	Alg	Rat	Gra	Fra	Mea	Dec	Int

The Element Dimension

Chapter 3 has shown that the nature of the elements (or the way they are interpreted) has a fundamental effect on task difficulty in generalised arithmetic. The elements in mathematics

*These were mainly developed by K Hart, D Küchemann, M Brown, G Ruddock and D Kerslake, and much of the statistical work was carried out by M McCartney. Chapters about each test can be found in CSMS (1980).

are commonly numerical, and here Collis(1975b) has shown, for example, that a change from small to large numbers can have marked effect on facility, which is confirmed by results from the CSMS Number Operations test (Brown and Küchemann,1976). A change from positive numbers to negatives, from whole numbers to fractions or decimals, and ofcourse from given numbers to unknowns, can also affect facility, as can be seen from the CSMS items shown below.

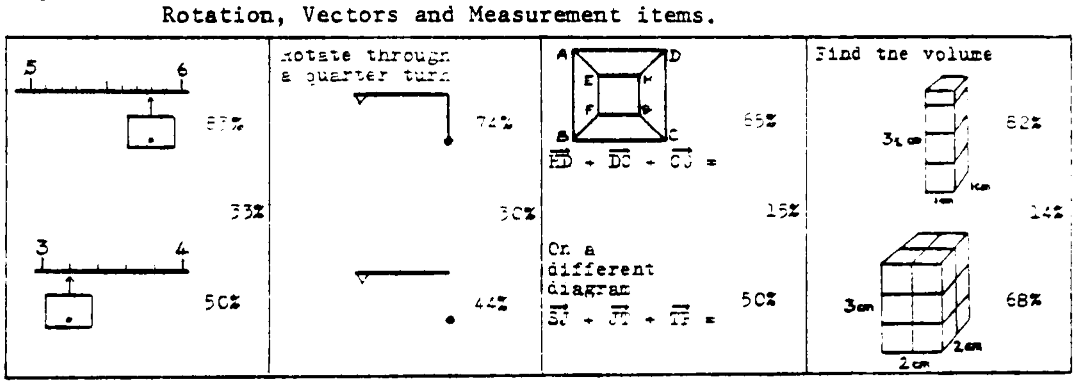
Fig 13.2 CSMS items showing the effect on facility of changes in the numerical elements.

		Facility (3rd year samples)	Change in Facility
small numbers to large numbers :	write a story for 9 x 3	45%	14% { [illegible]
	84 x 28	31%	
whole numbers to fractions :	plot (2,5)(3,7)(5,11)	91%	14%
	$(1\frac{1}{2},4)$	77%	
	volume of drawn cuboid, of dimensions: 3 x 2 x 2	68%	40%
	$2 \times 2\frac{1}{2} \times 2\frac{1}{2}$	28%	
	eels: x 3	75%	25%
	$\times 1\frac{1}{2}$	50%	20%
	x 5/3	30%	
	area of drawn rectangle of dimensions: 6 x 10	89%	76%
	2/9 x 5/8	13%	
whole numbers to decimals :	work out: 60 ÷ 3	80%	52%
	60 ÷ 0.3	28%	
	10 x 4	90%	32%
	10 x 5.13	58%	
	100 x 317	58%	22%
	100 x 2.3	36%	
	what does the 2 stand for: 521	87%	14%
	0.2	73%	9%
	0.260	64%	11%
	0.412	53%	
positive numbers to negative numbers :	work out: $^{+}2 + {}^{+}6$	97%	10%
	$^{+}8 + {}^{-}4$	87%	
	$^{+}6 - {}^{+}8$	70%	34%
	$^{-}6 - {}^{+}3$	36%	
	given route on a diagram, express as the vector: (2 1)	76%	13%
	$(1\ {}^{-}1)$	63%	9%
	$(0\ {}^{-}3)$	54%	
known numbers to unknown numbers :	number of sides → number of diagonals from vertex of polygon: 57→	75%	23%
	k→	52%	
	area of drawn rectangle of dimensions: 6 and 10	89%	77%
	5 and e+2	12%	

Other Dimensions

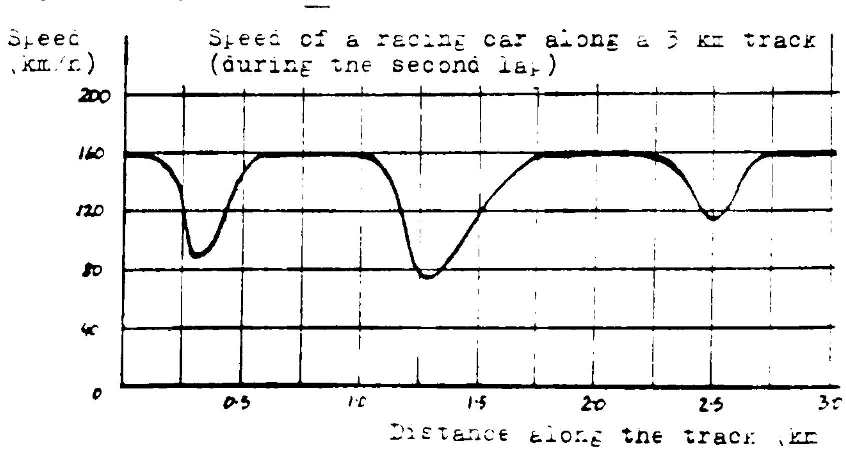
The above changes can in each case be described as a shift towards numerical elements that are less "intuitable". Sometimes such changes occur not so much in the elements themselves as in the way they are represented (see eg Malpas and Brown,1974, who distinguish between concrete and formal "models"). The first pair of items shown below both involve tenths, but in one case these can just be counted whilst in the other they have to be constructed (here 33% of 3rd years gave the answer 3.1); similarly a line has to be constructed from the second flag to the centre of rotation, whilst in the second Vectors item there is no supporting diagram at all, and in the case of the second cuboid some of the cubes are hidden.

Fig 13.3 3rd year facility of pairs of CSMS Decimals, Reflection and Rotation, Vectors and Measurement items.



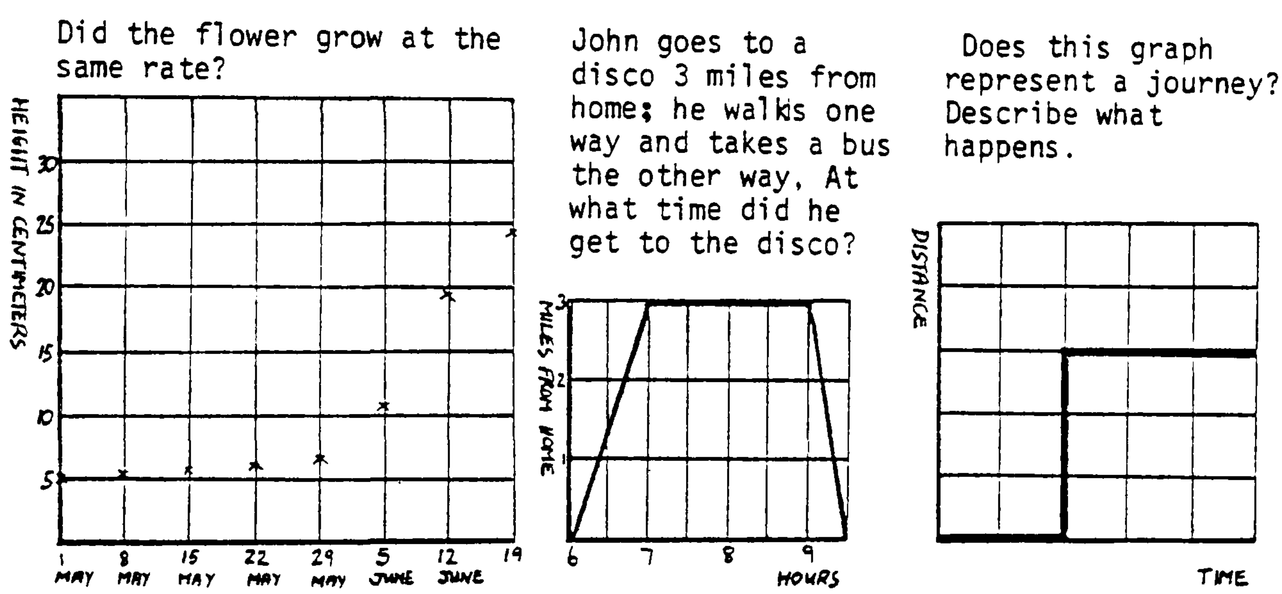
Mathematical representations may also cause difficulties because they are misleading. For example, Janvier (1978) found that a substantial number of secondary school children believed that the racing track corresponding to the graph below had nine bends rather than being triangular in shape.

Fig 13.4 Speed of a racing car along a 3 km track (Janvier).



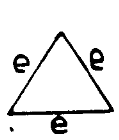
In the first of the three CSMS Graphs items shown below there is a direct correspondence between the graph and the situation it models: the horizontal axis corresponds to the ground and the height of the crosses corresponds to the height of the flowers. The second graph is somewhat misleading (John does not travel up a hill, along a plateau and down again) but it is possible to establish a correspondence between the three sections of the graph and the information given about John's activities. However, in the third graph no such information is given (here it should be said the item did not correlate particularly well, so that there might be other reasons for its low facility).

Fig 13.5 3rd year facility of three Graphs items.



97% facility   65% facility   8% facility

Though it may be stretching a point to regard the letters in generalised arithmetic as representations (rather than simply as elements) these can also be misleading. Thus, whilst the first item below can be easily solved by thinking of the letters as objects, children tend to see the letters in the second item in the same way, so that 4c+3b is interpreted as "4 cakes and 3 buns" and not as a number of pence.



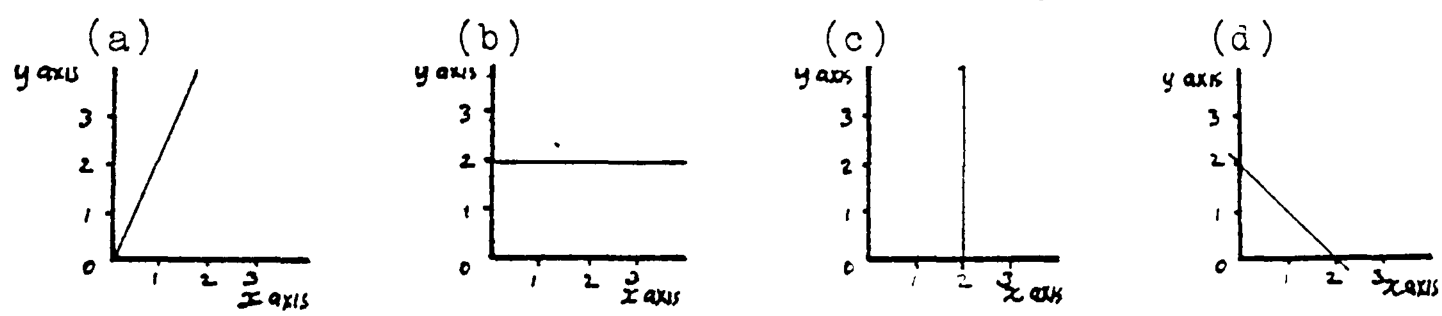
... (94%)

Cakes cost c pence each and buns cost b pence each. If I buy 4 cakes and 3 buns, what does 4c + 3b stand for? (22% facility)

Representations may sometimes be too "compact": this is particularly true of algebra, where letters may represent not not just isolated unknowns but sets of unknowns, sometimes also with a relationship between them. The same thing occurs with graphs: thus whilst 91% of 3rd years could plot the individual points represented by (2,5), (3,7) and (5,11), where only two pieces of information have to be related at a time, only 18% recognised that the first graph below, which can be said to summarise a whole set of relations, represents the line y=2x.

Fig 13.6 Graphs item.

Which of these graphs represents the line y=2x ?



Another important dimension concerns the operations involved in an item, of which one aspect is the type of operation, and another whether the operation is given or has to be determined:

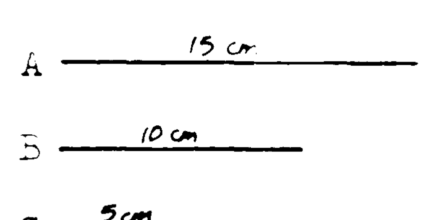
Fig 13.7 CSMS items where the operation is changed, or is left implicit.

Type of operation:	write a story for	84 - 28	77%	46%
		34 x 28	31%	
Explicit or implicit:		5/8 x 2/3	47%	
	1 km is the same as 5/8 miles; how long is 2/3 km in miles?		8%	39%

When several operations are involved, they sometimes have to be coordinated. For the Algebra item "Add 4 onto 3n" (36% facility) the operation +4 can simply be attached to what is given, whereas this would lead to an ambiguous answer in the case of "Multiply n+5 by 4" (17% facility). On the Ratio test a distinction can be made between items like the one shown below, where it is

Fig 13.8 Ratio eels item.

There are three eels, A, B and C.
They are fed sprats, the number
depending on their length.
If C eats 2 sprats, how many
should A be fed to match?



possible to build up to the answer (5:2 to 15:? via 10:4), and items where an appropriate correspondence first has to be constructed (eg 15:9 to 25:? via 5:3). A similar distinction arises on the Reflection and Rotation test, between reflections where the image can be constructed in a naturally ordered sequence of steps and those where the object has first to be analysed into discrete entities.

It is not always possible to keep the element, representation and operation dimensions entirely separate. Just as the letters in algebra might be thought of as representations rather than elements, so, for example, fractions can be thought of elements in their own right or as the result of operating on whole numbers; in ratio the key elements are operations; sometimes the nature of the operations determines how the elements can be interpreted: 2a+5b+a can be simplified by thinking of the elements as objects, but such an interpretation becomes strained for 3a-b+a.

Eventually it may be possible to reduce these dimensions to the single construct "processing capacity" (see the previous chapter), but it is probably more useful at present to build up a list of separate dimensions even if they sometimes overlap.

Similarities between Items at the same Facility Level

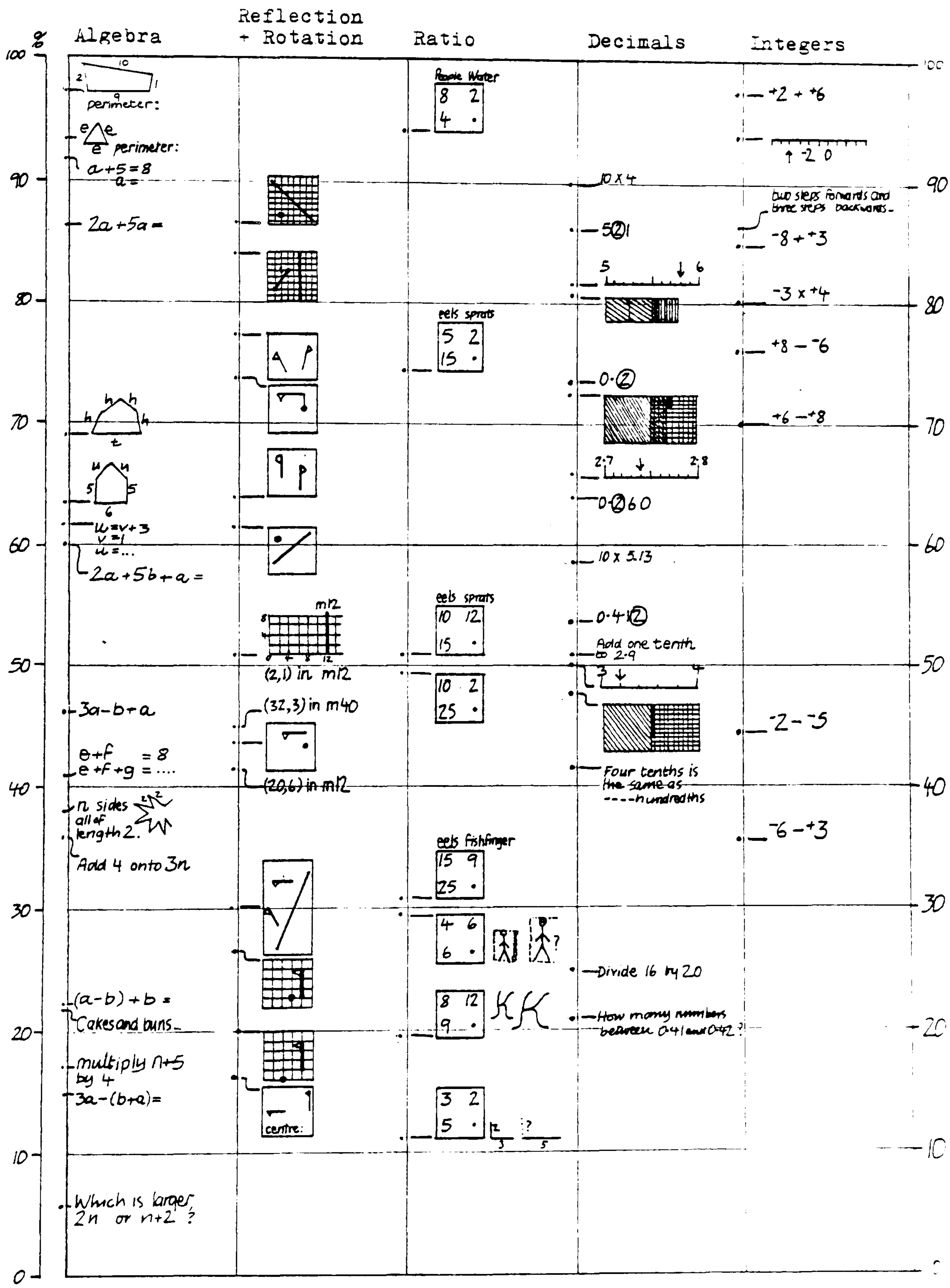
So far in this chapter differences in facility have been examined in one test and compared with differences in another test. It is now proposed to compare items from different tests more directly, by looking at items at given facility levels. As was suggested in Chapter 12, such comparisons are much more speculative and difficult to make than comparisons of facility differences, since there could be any number of reasons why an item has a specific facility. However some useful similarities do seem to exist, perhaps because an attempt was made to minimise, or at least to control, the complexity of the CSMS items.

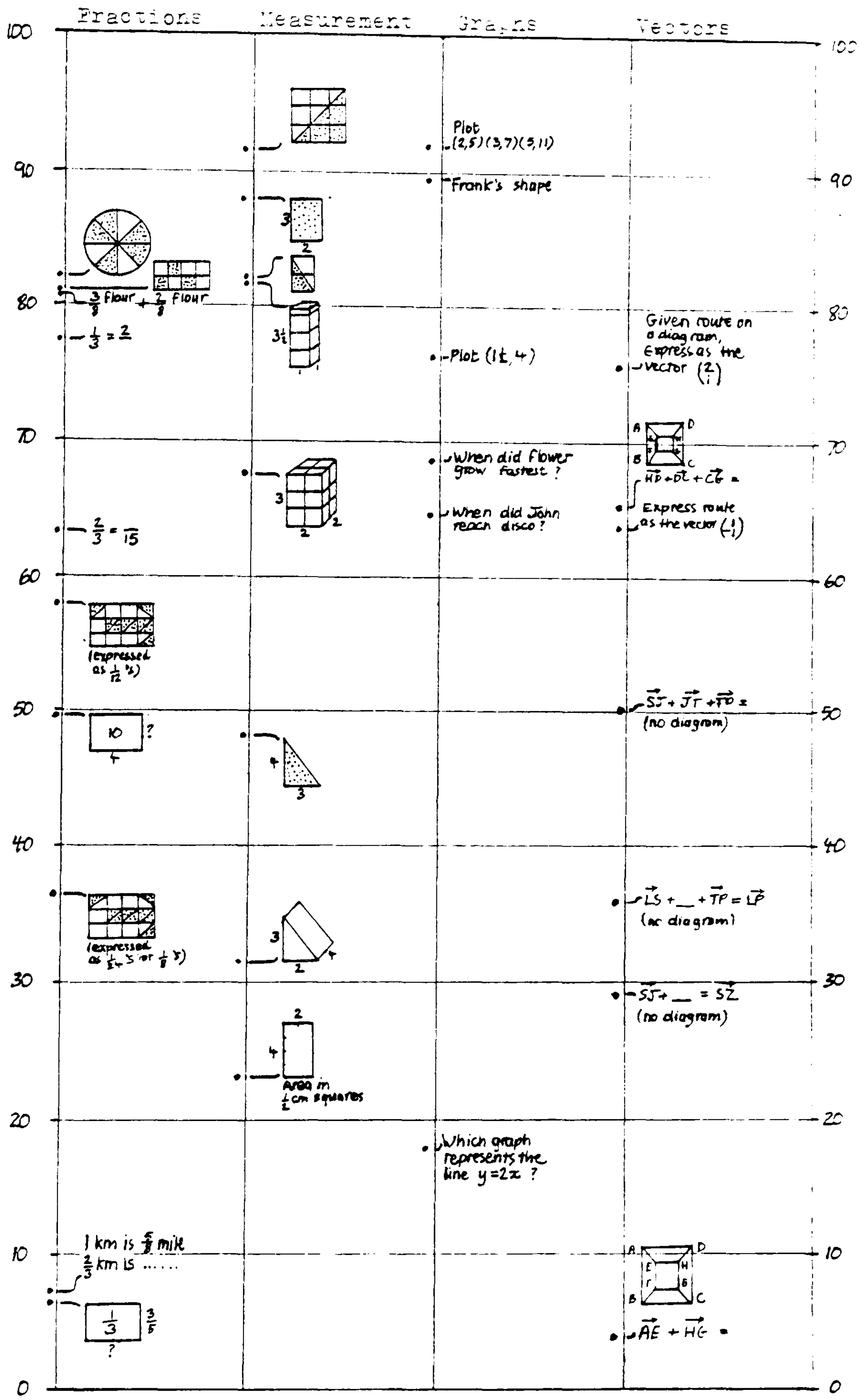
The items to be discussed are shown on the next two pages. Specifically, it is proposed to discuss the similarities between items at about the 75%, 50% and 25% facility levels, which represent the beginnings of levels 2, 3 and 4 on the Algebra test. Children will be described as being "at" a given facility level, in other words it is assumed that their performance within and across tests would be such that if they succeeded on some items at a given facility level they would succeed on most other items of the same and higher facility; though the facilities are derived from different (3rd year) samples, it can be argued that the samples are sufficiently well matched and the correlations within and across tests are sufficiently high for the assumption to be acceptable, at least as a working hypothesis.

The 75% facility level

Moving from the easiest items on the tests down to about the 75% facility level, it is possible to observe a shift from items involving whole numbers to items whose elements are formed by coordinating whole numbers in some way. Thus children who are able to cope with items at about the 75% facility level can use fractions (and decimals down to tenths) to represent shaded areas or subdivisions of a line, coordinates to represent points on a graph, and vectors to represent routes or ordered lists. In simple cases they can also perform reflections where more than

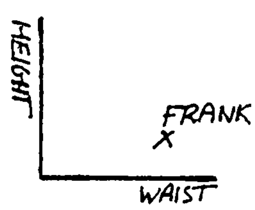
Fig 13.9 3rd year facility of selected items from CSMS mathematics tests.





Fractions
Measurement
Graphs
Vectors
100
90
80
70
60
50
40
30
20
10
0
Plot (2,5)(3,7)(5,11)
Frank's shape
3/8 flour + 2/8 flour
1/3 = 2
Plot (1½, 4)
Given route on a diagram, Express as the vector (2 1)
When did flower grow fastest?
When did John reach disco?
HD + DC + CG =
Express route as the vector (1 -1)
2/3 = 15
(expressed as 1/12's)
SJ + JT + TO =
(no diagram)
(expressed as 1/24's or 1/8's)
LS + __ + TP = LP
(no diagram)
SJ + __ = SZ
(no diagram)
Area in ½ cm squares
Which graph represents the line y = 2x ?
1 km is 5/8 mile
2/3 km is ......
AE + HG =

one aspect (distance and direction) has to be controlled in sequence, determine areas by counting fractional parts of squares and volumes by counting cubes (when these are all visible). These children can interpret isolated features of graphs when these correspond directly to reality, as in the case of Frank, below, who is recognised as short and fat when the axes are oriented as in the diagram, but not when the axes are interchanged. Children at this level are also starting to relate such intutiable features: for example, for the graph showing the weekly height of a flower they can recognise when it grew fastest.



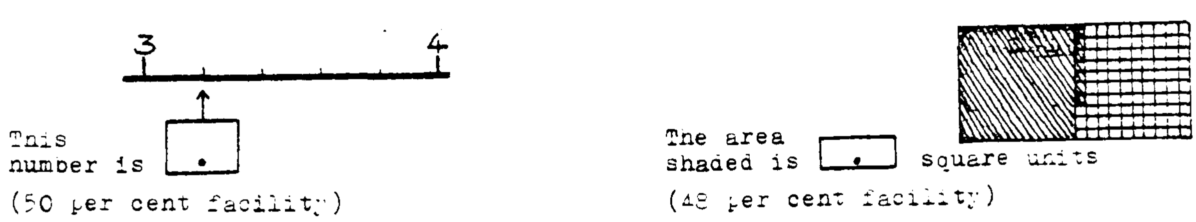
In certain circumstances children at this level can also operate on these new elements; in one of the eels items on the Ratio test they can construct the correspondence 15:6 from 5:2 (A 5 cm eel eats two sprats; how many does a 15 cm eel need?); they know that 2/6 is equivalent to 1/3 and they can correctly add 3/8 and 2/8 in a realistic context (without which they still tend to give the answer 5/16); they can also add integers by coordinating shifts along the number line, and even solve simple subtractions like $^{+}6-^{+}8$ (instead of just taking the smaller number from the larger). However, these new elements are not yet integrated into a system: in ratio they cannot yet construct correspondences that require more than repeated addition, nor can they establish the equivalence of less familiar fractions, nor differentiate tenths from other decimals (they recognise that 2 in 0.2 represents tenths, but in 0.250 they may well say it stands for hundreds or hundredths).

The 50% facility level

As children begin to cope with items beyond the 75% facility level, their understanding of these new elements is extended: they begin to differentiate and impose an ordering on progressively less intuitable decimals; the can cope with route vectors involving negatives and zero; they can count hidden cubes and they begin to accept unclosed answers like 4h+t (instead of 4ht or 5ht) and 3a+5b (instead of 8ab), but only when the letters can be regarded as objects. However, it is not until children are at about the 50% facility level that a definite shift away from directly intuitable reality can be observed.

At about the 50% level, children can apply appropriate transformations to the decimal representations shown below, instead of simply recording entities smaller than 1 in the first available decimal place (35% and 32% of 3rd years gave the answers 3.1 and 1.7); children can now find a vector equivalent

Fig 13.10 3rd year facility of two Decimals items.

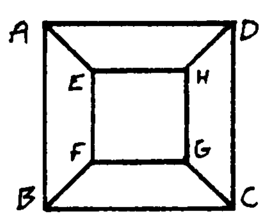


to $\overrightarrow{SJ} + \overrightarrow{JT} + \overrightarrow{TP}$ without a diagram, and describe the effect of the translation $\begin{pmatrix}9\\8\end{pmatrix}$ followed by $\begin{pmatrix}-9\\-8\end{pmatrix}$, where the original translation goes off the page; with an appropriate diagram, there is also a marked increase (Ruddock,1980) at this level in the proportion of children who choose to combine translations by adding vectors rather than counting squares. These children are also beginning to cope with reflections that go off the page and with rotations where the object does not pass through the centre of rotation. They can determine the area of a triangle by going beyond the confines of the figure (ie by constructing a rectangle), and they are beginning to use letters as unknown numbers. In ratio children still tend to build up their answers by repeated addition, but they can now construct mediating correspondences arising from this method (eg 5:1, in order to get from 10:2 via 20:4 to 25:?).

The 25% facility level

Children at this level have even less need of the support provided by familiar elements or representations that correspond directly to reality. They can not only find the missing vector in the naturally ordered sequence $\vec{SJ} + \vec{JT} + \vec{TP} =$ __ without the aid of a diagram, but in $\vec{LS} +$ __ $+ \vec{TP} = \vec{LP}$ and $\vec{SJ} +$ __ $= \vec{SZ}$. The can perform reflections by analysing the object into abstract elements, rather than by building up the answer in a step-by-step manner. Similarly, they can construct mediating correspondences in ratio when the building up method no longer applies (eg 5:3 for 15:9 to 25:?). They can subtract integers and divide decimals and are no longer tied to the intuition that these operations always produce smaller answers; nor do they assume that halving linear measurements halves the area. In Algebra they not only accept that a letter may represent a range of values but that for different letters these (unknown) values may coincide.

On the other hand, it is not until children can cope with items beyond this level that they seem able to work with abstract mathematical systems per se. Then they can determine the side of a rectangle of area 1/3 whose other side is of length 3/5, and enlarge a figure by a factor 5/3 with no mediating correspondence of whole numbers to help them (3:2 to 5:?). In Algebra they can then operate on operations (eg simplify $3a-(b+a)$; multiply $n+5$ by 4) and establish second-order relationships by comparing not just unknowns but the changes in the value of one with the changes in another; also, in items involving numerical relationships between objects they no longer succumb to using the letters as a shorthand for the objects themselves. In Graphs they are no longer restricted to interpreting isolated features but recognise that a graph may represent the relationship between an entire set of ordered pairs. In Vectors their understanding has become sufficiently flexible to accept the notion that seemingly unrelated vectors like $\overrightarrow{AE}$ and $\overrightarrow{HG}$ in the diagram below can be combined.



Of the features characterising each facility level, the ones at the 50% level are perhaps the most coherent across tests, and it is tempting to see this level as signalling the onset or transition to formal operations. In practical terms this level is probably also the most interesting as it is the one that the majority of secondary school children will just achieve or spend their time struggling towards.

Chapter 14 RELATED STUDIES AND SOME SUGGESTIONS FOR FURTHER RESEARCH

In this chapter it is proposed to discuss a small selection of studies that have a fairly direct bearing on the CSMS Algebra research. In the course of this discussion a number of suggestions will made of how the Algebra items could be extended in order to produce items that might throw further light on the interpretation of children's understanding of algebra put forward in this thesis.

There are ofcourse many directions that research based on the present study could take, and before embarking on the discussion outlined above three of these will be mentioned briefly. Probably the most obvious next step, whether or not further items are constructed, is to go back to interviewing individual children, bearing in mind that the interpretations presented in this thesis are based primarily on children's written responses (albeit incorrect as well as correct ones). Infact interviews using items from the Algebra test are currently being undertaken by Lesley Booth as part of the "Strategies and Errors" project at Chelsea College.

A second direction in which future research might go is to look in detail at a specific area of secondary school algebra, such as "equation solving", "substitution", "manipulation", etc. There are items on the Algebra test under each of these headings but they can clearly not be investigated adequately in a single test (a point made in Chapter 2). Work currently being carried out at Nottingham University by Alan Bell and Christine Shiu is very much along these lines and is looking, amoungst other things, at secondary school children's methods of solving equations. An interesting aspect of this work is the finding that the methods children use (and the meanings given to the unknowns) depend very much on the nature of the equations they are being asked to solve.

A third direction, and one that is a feature of both the Chelsea and the Nottingham projects, is to examine the effectiveness of various ways of teaching algebra. Such research is of obvious practical interest and it is also important theoretically since it may throw light on the question of the extent to which children's understanding is enhanced by certain experiences at school and constrained by cognitive developmental factors. The effect of teaching can also be examined in a more global way, by for example comparing the performance of children from different year groups or from different classes or schools. Here some data already exist (see, for example, Chapter 6), though the conclusions that have emerged to date are somewhat ambiguous (perhaps because the data have not been analysed in sufficient detail, but also because some of the information about children's schooling and other background variables is not very precise).

Returning to one of the main purposes of this chapter, the value generally of extending a set of items in a written test (as well as going back to individual interviews) can be illustrated with reference to the items below, taken from Firth (1975).

Fig 14.1 Equation-manipulation items, and their facilities (Firth).

	$P = R + S - T$	Facility (15 year olds)
a.	$P - R =$	50
b.	$P + T =$	76
c.	$P - S + T =$	65
d.	$S - T =$	29
e.	$T - S =$	35

The items were designed to test children's knowledge of simple rules for manipulating equations. At first sight they all look very much the same, and it might be asked why 5 items rather than just 1 or 2 were used. However, an examination of their facilities (obtained from a 4th year class at about the O level / CSE borderline) shows clearly that they are not all the same; in particular, there is a substantial difference in facility between items where P remains on the left hand side of the equation (a, b, c) and where it does not (d, e), and between items where either R (item a) or T (item b) are moved over to the left hand side. Bell (in Bell, Costello and Küchemann,1980) puts forward a number of possible explanations for these facility differences. He then states

> "Whether or not these explanations apply, the clear differences between the scores certainly show that we are dealing with something other than simply knowing or not knowing a simple manipulative rule. There are factors affecting the understanding and use of these rules which often go unrecognised and therefore untreated in the course of teaching" (ibid, p112).

Similarly there may be a multiplicity of factors affecting children's understanding in other areas of algebra, and clearly this is as much a problem for research as it is for teaching.

Bell suggests that the difference in facility between items a, b, c and d, e may be due to children not seeing the equals sign as fully reversible. There are other studies that lend support to this view (eg Lunzer, Bell and Shiu,1976; Kieran,1980). However, it can also be argued that to solve items a, b and c children only need to cancel from the right hand side the letter that has been moved over to the left, whilst for d and e they not only have to determine which letter has been moved but to make a decision about its sign. To determine which, if any, of these explanations holds, children could be observed solving the items and asked about the strategies they used, or further items could be written, such as the ones below in which the original equation has been reversed. Here, if the first explanation holds the facilities should be reversed for a, b, c and d, e; if the second holds they should stay more or less the same.

Fig 14.2 Variants of Firth's items.

$R + S - T = P$

a. $P - R =$
b. $P + T =$
c. $P - S + T =$
d. $S - T =$
e. $T - S =$

Firth also used an item in his study which asked children to write down a number "which is 3 more than x". About a third of the pupils could only cope with this by giving a value to x (letter evaluated) rather than using x as a specific unknown, and on interview such children made statements like "I can't do this until you tell me what x is". Similar responses occurred with the original version of question 15 on the Algebra test when children were asked for the number of diagonals that can be drawn from a vertex of a k-sided polygon (even though they might be able to express the rule in general terms such as "You take away 3 from the number of sides" -which raises the question, to be taken up again in the next chapter, of whether children's difficulties stem at least in part from their being asked to do something which is not so much incomprehensible as unhallowed and pointless). In the APU primary survey (APU,1980a) about half the 11 year olds could find the 100th term of "simple number patters" but "almost no-one was able to give the nth term" (ibid, p53).

Ekenstam and Nilsson (1979) also allude to the difficulty of determining the factors affecting children's understanding of a mathematical task:

> "Most problems are difficult to analyse because of their complexity. There are many possibilities for the students to make mistakes with a certain problem and it is not very interesting to ascertain that a certain percentage of the students have given a wrong answer to a problem. Look for example at the problem
>
> Express $(3z + 1)^2$ without brackets.
>
> A pupil who fails may not understand
> - the meaning of the exponent 2,
> - that the square of 3z is $9z^2$ but answers $3z^2$,
> - that the expression has three terms and forgets 6z,
> - how to multiply $(3z+1)(3z+1)$,
> - that $z \times z = z^2$ but believes that $z \times z = 2z$,
>
> many other possibilities not mentioned" (ibid, p44).

Ekenstam and Nilsson tackle this problem in a most elegant way, by breaking down a task into a series of simpler tasks (assumed to represent the steps required to solve the task) and then giving the simpler tasks (with variations) to different but equivalent samples of children. In this way they tested 130 items (in algebra and geometry), using a total sample of about 2000 children, with each item being given to about 200 children (Swedish 16 year olds representing roughly the top 20% of the ability range). One such set of items is shown below, together with their facilities.

Fig 14.3 Set of item-variants, and their facilities (Ekenstam and Nilsson).

Test item		A %	Content
(1)	$\frac{3x-2}{2} = \frac{x}{3}$	28	
(2)	$3(3x-2) = 2x$	70	The first step performed
(3)	$9x-6 = 2x$	74	The second one performed
(4)	$9R-6 = 2R$	64	As (3) but x is changed to R
(5)	$7x-6 = 0$	71	A possible step after (3)
(6)	$7z-6 = 0$	72	As (5) but x is changed to z
(7)	$7x = 6$	77	Final step
(8)	$243x = 242$	69	As (7) but the coefficients are changed

One drawback with their method is the use of different subsamples (even though they are quite large), and this may be the reason why, for example, replacing x by another letter in one case (items 3 and 4) appears to produce a substantial change in facility but in another case (items 5 and 6) produces hardly any change at all. (According to the APU primary survey, the change from geometrical symbols to the use of letters to respresent unknowns in simple equations had a substantial effect on facility; also, Wagner,1979, states that children tend to believe that different letters used in the same context represent different values. However, neither of these studies indicate that changing from one letter to another necessarily makes a task more difficult.) Despite the above drawback, Ekenstam and Nilsson's study produces some interesting comparisons, as for example with the three equations shown below (in which children were asked to solve for x or to make t the subject of the equation). In terms of the Algebra research, the first item is easier because the letter can be evaluated (albeit not as a positive whole number) whereas in the other two items the letters have to be treated at least as specific unknowns.

Fig 14.4 Facilities of items involving letter evaluated and letter as specific unknown (Ekenstam and Nilsson).

Item	Facility
$4 = 3x$	70
$v = 3t$	54
$v = at$	44

The following three items, though they all require that the letter be evaluated, bring out the important difference between the mathematical structure of an item (in terms of which item b would appear to be the msot difficult) and an item's psychological structure (items a and b involve only positive whole numbers and can be solved intuitively or by trial and error, whereas c is difficult to solve other than formally -multiplying both sides by x, etc).

Fig 14.5 **Facilities of items involving whole numbers and fractions (Ekenstam and Nilsson).**

Item	Facility
a. $\frac{30}{x} = 6$	82
b. $\frac{14}{x+2} = 2$	58
c. $\frac{4}{x} = 3$	48

On the Algebra test there are many items for which it would be interesting to construct and test variants along the lines of Ekenstam and Nilsson's study. For example, with respect to items 13vii and 13ix (Simplify $a+4+a-4$ and $(a+b)+(a-b)$) intermediate items like $(a+4)+(a-4)$ and $a+b+a-b$ could be used to investigate whether the drop in facility (from 44% to 19% for the 3rd year sample) is primarily due to the insertion of brackets in 13ix or to the substitution of a number (4) by a letter (b). Similarly it would be of interest to compare item 13v (Simplify $(a-b)+b$) with $a-b+b$ (which, incidentally, Keats,1955, classifies as an item requiring formal thought) and, further, to compare this with $a+b-b$, in order to test the assumption that $a-b$ requires a greater acceptance of lack of closure than $a+b$ (particularly if the letters are regarded as objects). Ofcourse this can be extended further still, for example via items like $a-7+7$ and $4-7+7$ to, say, $8-7+7$ (which Keats classifies as concrete).

With younger children it might be interesting to compare items like 6i (What can you say about a if $a+5=8$) with variants such as $\square+5=8$, $8=a+5$ and $5+a=8$. Studies mentioned earlier in this chapter suggest that the first two variants would both be easier than the original item and evidence cited by Bruner suggests that this is also true of the third variant -perhaps because of "the transfer of linguistic habits from ordinary English, where sentences are easier to complete when a term is deleted from the middle than from the beginning of the sentence" (Bruner,1966,p55).

Linguistic habits may also play an important rôle in some of those very difficult Algebra items where children were asked to translate a mathematical expression into English, or vice versa. In these items, it will be recalled that there was a strong tendency to interpret the letters as objects (cabbages and turnips, blue and red pencils, etc) rather than numbers of objects. In part this may have been due to the fact that in all these items children were asked to use "initial" letters to represent the numbers of objects, rather than "neutral" letters like x and y. This issue was investigated during the development of the test, but not adequately. For example, question 20 (where the expression 4c+3b represents 4 cakes at c pence each and 3 buns at b pence each) appeared on Draft 4 in a parallel form to question 10 (where 8c+6t represents c cabbages costing 8 pence each and t turnips costing 6 pence each) but with the letters x and y instead of c and b (4x+3y representing x cakes at 4 pence each and y buns at 3 pence each). Unfortunately both questions proved to be too difficult for all the children in three of the classes that were tested (N=69) whilst just 2 children answered both items correctly in another class (N=21). It was concluded that the use of neutral letters made no difference but clearly this was premature and the questions should have been tried with children of higher ability. On the other hand, it is also clear that the use of neutral letters does not eliminate the tendency to interpret the letters as objects. This is confirmed by Galvin and Bell (1977) who found, when working with variants of item 22 (blue and red pencils), that even subjects who used the letters x and y and defined them correctly sometimes reinterpreted expressions like 6x as 6 red pencils.

One of the variants used by Galvin and Bell was as follows:

Fig 14.6 Variant of Algebra item 22 (Galvin and Bell).

a. Tom buys 1 red pencil and 6 blue pencils. He pays 23 pence altogether.
b. Jane buys 3 red pencils and 7 blue pencils. She pays 36 pence altogether.
c. Find the cost of 1 red pencil and the cost of 1 blue pencil.

Part a can be modelled by $r+6b=23$ (or $x+6y=23$, etc), and pupils who attempted to solve the item by constructing a pair of equations of this sort found this aspect of the task remarkably easy. However, upon being asked about the meaning of their equations it became apparent that many of the pupils were not interpreting them mathematically (ie as pure numerical relationships) but instead were regarding them simply as a symbolic shorthand for the original English sentences:

1 red pencil	and	6 blue pencils	cost	23 pence
1r	+	6b	=	23.

Another example cited by Galvin and Bell is of a pupil who interprets the sentence "If one rabbit is put in each hutch, one rabbit will be left without a place" as "There is one more rabbit than there is hutches" and who then writes $r+1=h$. However, this time there is no spurious match between the symbolic shorthand and the mathematical statement $r=1+h$ which describes the relationship between the <u>number</u> of rabbits and hutches.

Given the apparent tendency of pupils to construct symbolic expressions where the order of the symbols corresponds directly to the phrase order of the original sentences, Galvin and Bell suggest that pupils should be encouraged to re-word these sentences in terms of the defined unknowns: for example, the statement about the rabbits and hutches should be transformed into "The number of rabbits is one more than the number of hutches". Whether this would lead to correctly formed and correctly interpreted mathematical relationships is clearly worth investigating, as is the question of how easily children would be able to make the initial verbal transformations.

Returning for a moment to item 22 on the Algebra test (blue and red pencils), a comparison of this (and the rabbits and hutches) item with the variant used by Galvin and Bell suggests there may be a crucial difference between using a letter to represent the number, per se, of a set of objects and using it to represent a numerical property of a set of objects (such as their common length, weight or, in the case of Galvin and Bell's variant, cost). This is supported by findings from the APU secondary survey (APU,1980b) where an item in which children were asked to write an expression for the length of an n-rung ladder had a facility of under 10%, compared to 55% for the item shown below:

> Write an expression for the total cost of three bars of chocolate and a packet of crisps, when a bar of chocolate costs x pence and a packet of crisps y pence.

(Ofcourse, as Galvin and Bell's study has shown, the fact that 55% of the children were able to derive the correct expression does not mean that they all interpreted the expression in a mathematically correct way.)

Studies by Clement and his associates suggest that the use of letters as objects in translation tasks such as the ones investigated by Galvin and Bell persists well beyond school age. For example, item 1 below was answered correctly by only 63% of a sample of 150 engineering students at two American universities, with 25% giving the "reversal" (or letter as object) response 6S=P (Clement et al,1980a). Item 2 was given to 17 engineers, each with at least 10 years professional experience, of whom only just over half (9 subjects) gave a correct response (Clement et al,1980b).

Fig 14.7 Translation items (Clement).

Item 1 (Students and Professors)

Write an equation using the variables S and P to represent the following statement: "There are six times as many students as professors at this University." Use S for the number of students and P for the number of professors.

Item 2

Given the following statement: "At the last football game, for every 4 people who bought sandwiches, there were 5 who bought hamburgers.", write an equation which represents the above statement. Use S for the number of people who bought sandwiches, and H for the number of people who bought hamburgers.

Rosnick and Clement (1980) devised a written teaching unit, part of which is shown below, that was given to 6 students who had made the reversal error on the Students and Professors item (item 1).

Fig 14.8 Teaching unit for translation items (Rosnick and Clement).

Step 1. Understand the English sentence and describe what is asked for in your own words. Find numbers that would fit the relationship.

Step 2. Attempt to write an equation.

Step 3. CHECK YOUR ANSWER in the following way: REPLACE the letters in your equation with the numbers you found in Step 1 and see if both sides of the equation really are equal. If not, repeat Step 2.

The unit was highly effective, in the sense that all the students were eventually able to produce correct equations for 4 items similar to items 1 and 2 above. However, Rosnick and Clement argue that for at least 5 of the 6 students there was no real change* in their understanding. For example, after having written a correct equation one student commented

> "This is probably right because it works. It works (by plugging in values) but I don't know why it works... It works but I don't think it works."(ibid,p17).

Another student, on returning to her correct solution to the Students and Professors item (S=6P), said

> "For every student there are.. no, for every.. see, it's not for every student there are 6 professors.. I don't know. I'm confused now."(ibid,p17).

Clement argues that there are two distinct strategies leading to the reversal error (eg 6S=P instead of S=6P). One he calls "word order matching" or "syntactic translation" (cf Galvin and Bell,1977), whereby a direct mapping is effected from the verbal to the symbolic expression without the given item being interpreted in any meaningful way, whereas in the second strategy students demonstrate a proper understanding of the situation described in the item but the letters are then used as objects:

> "To these students the letter P stands for 'a professor' rather than 'the number of professors'"(Clement et al,1980b,p6).

It can be argued that for simple items like Students and Professors the distinction between these two strategies is not very real: it seems more likely that such items are interpreted meaningfully but that the letters are used as objects under both strategies. However, with respect to the second strategy at least,

*It can be argued that the very confusion shown by these students indicates that <u>some</u> advance in understanding has been made.

Clement's further comments are illuminating: as well as not using the letters to represent numbers of objects, Clement suggests that students who produce answers like 6S=P are describing "a passive picture in which relative sizes of the entities are represented" instead of expressing "an active operation being performed on one number (the number of professors) in order to obtain another number (the number of students)" (Clement et al,1980b,p7). Put more succinctly, their descriptions are expressed as relations rather than operations. The distinction between these two kinds of description may be just as important as that between the use of letters as objects and as numbers of objects, when it comes to translating verbal statements into an algebraic form. Consider the two statements below (of which the second was given by a student who correctly wrote S=6P; Clement et al,1980b,p7):

> The number of students equals six times the number of professors.
>
> If you want to even out the number of students to the number of professors, you'd have to have six times as many professors.

Though both statements are phrased in terms of the required numerical unknowns, it seems likely that the first one, which describes a relation rather than an operation (and which is of the form suggested by Galvin and Bell) is more readily mistranslated than the second; unfortunately, the second statement appears to be less natural (it is not a direct description of the existing state of affairs) and is therefore likely to be more difficult to construct. (Operational statements are required in computer programming, and it is interesting to note that subjects were significantly more successful at producing correct [illegible]s in this context; Clement et al, 1980b.)

Shiu (1978) has investigated the ways in which children solve items similar to 5i and 5ii on the Algebra test.

Fig 14.9 Algebra items 5i and 5ii.

Item 5i	Item 5ii
a + b = 43	n - 246 = 762
a + b + 2 = ...	n - 247 = ...

It will be recalled that these items were classified under the category letter not used, on the argument that they could be solved by matching. However, while Shiu found that mathematically competent 16 - 17 year olds did tend to solve such items in this way, 11 - 12 year olds were more likely to use what she calls "sequential closure" (ie letter evaluated), which suggests that this may be a more primitive, but still sometimes successful, strategy for items of this sort. Shiu goes on to argue that matching requires a greater degree of acceptance of lack of closure.

From children's written responses to items 5i and 5ii it was not usually possible to determine the method used, although on item 5ii in particular there were a few children whose working clearly indicated that they had used sequential closure -and ofcourse other children who did not show any working may also have used this strategy. As far as the wrong answers to item 5ii are concerned it seems likely that the children who gave 763 (16% of the total sample) used matching, whilst other wrong numerical answers (8%) were obtained by sequential closure. Interestingly, children who gave the answer 763 performed, on average, slightly better on the rest of the test than those who gave other wrong answers, which tends to support Shiu's suggestion that matching is a higher level strategy for items of this sort.

Draft 4.1 of the Algebra test contained these three items:

Fig 14.10 Variants of item 5ii from Draft 4.1.

n + 4 = 20	n + 30 = 6	n + 364 = 543
n + 5 = ...	n + 32 = ...	n + 365 = ...

These can each be solved by matching or by first evaluating n. However, while the matching strategy would appear to be of similar difficulty in each case (+1, +2, +1), it can be argued that evaluating n would lead to a greater number of errors in the second item (n negative) and in the third (large numbers leading to arithmetical errors) than in the first item. The items were given to an above average class of 13 year olds (N=29) and were each answered correctly by about the same number of children (20, 19 and 19 respectively), which suggest that most of this sample (and in particular most of those who were successful) used matching. It would be interesting to give these items to a larger sample, and also to younger and less able children to see whether a point is reached where most children do use sequential closure, as Shiu suggests, and if so, to see to what extent the strategy is used succesfully.

The items below were used in the APU surveys, where they are described as items where "the variable was not to be evaluated but acted as a sort of place holder" (APU,1980a,p53), though as with the final version of the Algebra test, it is not certain that children saw them in this way. The facilities seem to be comparable with those obtained on the Algebra test, bearing in mind the differences between the items and the age groups tested.

Fig 14.11 Variants of item 5ii, and their facilities (APU).

Item	11	15	age
			facility
n stands for a number. n + 4 makes 21 so n + 5 makes	-	90	
n + 4 = 21 so n + 5 = ...	63	-	
B - 9 = 21 so B -10 = ...	51	-	

An investigation by Kieran (1979) into children's knowledge of the conventional order of arithmetical operations and of the use of brackets is of interest in relation to item 4iii on the Algebra test (Multiply n+5 by 4). On this item a substantial proportion of children (19% of the 3rd and 18% of the 4th year samples) gave ambiguous answers like n+5x4, rather than say (n+5)4, and in Chapter 5 it was argued that such answers could not be explained solely in terms of a lack of familiarity with the appropriate conventions since most of these children would have met these conventions at some time in their school lives. An alternative explanation is that ambiguous answers arise from an inability to consider other possibilities (an unawareness, rather than a non-acceptance, of lack of closure); in other words, children who give ambiguous answers are at a lower cognitive level than those who recognise the ambiguity and therefore resolve it.

Kieran interviewed junior high school students, all of whom had been taught the appropriate conventions. However, she found that none of the students fully grasped the conventions and when asked to conceive or evaluate strings of operations the students simply worked in the order the operations were written, ie from left to right. These findings support the view that familiarity with the conventions is not enough. However, rather than stemming from cognitive-developmental factors, Kieran offers a third explanation for children's difficulties with the conventions, namely that they have not been presented with situations which demonstrate that conventions other than working from left to right are needed. This explanation can itself be countered by a cognitive-developmental argument (children have met such situations but have failed to grasp their significance). However, Kieran's explanation is clearly worth further investigation -for example by assessing the effect, on different children, of situations of the type Kieran suggests. One such situation might be to present a string like 3+4x5, which according to the left-right rule equals 35, and then to transform it in a way that is accepted as legitimate but which changes its value according to the rule -for example by re-writing the string as 3 + 2x2 x 5.

Apart from the seminal work of Collis, undoubtably the work most directly relevant to the Algebra research is that of Harper (1978,1979,1980), whose analysis of what he calls a "numerical variable" fills an important gap between what has here been called "variable" and the rather poorly articulated notion of "generalised number". Harper (1980) discusses the use of a letter as a "non-ordered numerical entity" (described on another occasion as a "pregnant numeral"), which perfectly captures the notion of an entity in its own right which at the same time is not a unique, albeit undetermined number ("specific unknown"). This notion is traced back to Vieta who in 1591 introduced the concept of what he called a "species".

Harper devised the following item to distinguish between children who, at best, see letters as specific unknowns and those who have the concept of a non-ordered numeral:

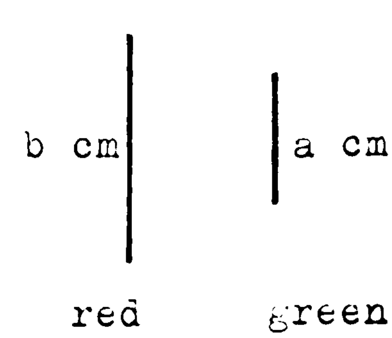
Fig 14.12 Item for testing notion of non-ordered numeral (Harper).

Is the red line longer
than the green line..?

Why?

When is the green line
longer than the red line?

etc



Responses were classified into two types: Type A, where the meaning of the letter was derived from viewing the two lines as "concrete" ("If you double or treble the green line it will be longer"; "If the green line is in the distance it might be longer"), and Type B, where the letter was given a meaning independently of, or by transcending, the "concrete" data ("The green line is longer when a is greater than b").

With one of his students at Bath (McLeay,1980), Harper designed parallel versions of an algebra test, of which one version (Test S) incorporated the geometrical distractor used in the above item, as well as other distractors (or "repeller factors" as McLeay calls them). The tests were each given to one of two samples of 26 14 year olds from the same comprehensive school, who had been carefully matched in terms of their level of performance and total score on a shortened version of the Algebra test. The strength of the geometrical distractor (and by implication, children's tenuous hold on the non-ordered numeral concept) is demonstrated by the two items below, which were answered correctly by 22 and 4 children respectively.

Fig 14.13 Pair of items for testing non-ordered numeral (Harper).

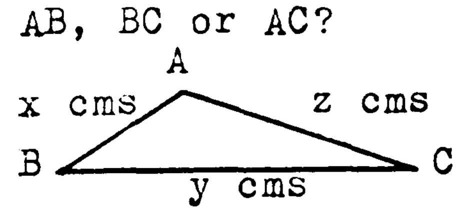
Item 3, Test A

Triangle ABC has sides AB=x cms, BC=y cms and AC=z cms.

Which side is the longest?

Item 3, Test S

Which side is the longest, AB, BC or AC?



Examples of the other distractor types used by McLeay are shown by the item-pairs below; the figures in brackets show the number of correct responses for each item (N=26). For each distractor type, McLeay computed the ratio of total score on the appropriate Test A items to total score on the corresponding Test B items. The ratio came to 2.7 for the geometric distractor, whilst for the numeric, diagramatic, verbal and symbolic distractors she obtained the values 2.2, 1.2, 1.8 and 1.7 respectively. This indicates that the geometric distractor was generally the most powerful, though the effect of the others was sometimes still substantial.

Fig 14.14 Items for testing effect of different distractor types on the notion of non-ordered numeral (McLeay).

Distractor Type	Item from Test A	Corresponding Item from Test B
Numeric	Item 1 (24) Which is larger, a or 12 ?	Item 1 (11) Which is larger, 14a or 12 ?
Diagrmtc	Item 17 (25) A zigzag line has n sides each of length p cms. What is the total length of the line making the zigzag?	Item 17 (21) Some of this zigzag is hidden. etc 
Verbal	Item 8 (23) If 1+m is larger than 1+t, what can you say about t?	Item 8 (4) 1+m is larger than 1+t always, sometimes, or never?
	Item 6 (9) 1+m = 1+t always, sometimes, or never?	Item 6 (4) Which is larger, 1+m or 1+t ?
Symbolic	Item 15 (10) Which is larger, a+b or a+c ?	Item 15 (3) Which is larger, a+b or a-c ?

It may have been noticed that some of these items have similarities with item 3 from the Algebra test (Which is larger, 2n or n+2 ?) and indeed the item appeared on Test B, where it was answered correctly by 4 children. In terms of the classification illustrated above it might be argued that the item suffers in particular from a symbolic distractor: where in item 15 (Test B) children have to overcome the belief that subtraction always makes things smaller, in the Algebra item 3 it is necessary to realise that multiplication is not always "stronger" than addition (a common response was to say 2n was larger than n+2 "because it's multiply"). However, this does not necessarily mean that item 3 is testing only for understanding of the non-ordered numeral concept, even though it seems to be of the same order of difficulty as item 15 on Test B: in item 3 the crucial value is n=2 whilst in item 15 children have to consider negatives, which in itself is likely to be more difficult; in turn this suggests that the task of actually finding the value n=2 requires something in addition to (or at least other than) the non-ordered numeral concept (namely some understanding of second-order relations).

Before becoming too enmeshed in trying to preserve a distinction between the constructs of variable and non-ordered numeral, it should be recalled that McLeay's data were derived from very small samples of children. More importantly perhaps, it is worth pointing out that though variable would appear to be a higher level concept than non-ordered numeral (to construct a second-order relation it is necessary to realise that the unknowns can take on a range of values) this does not mean that all items involving the notion of variable have to be more difficult than those involving only non-ordered numerals: not only is it possible to construct second-order relations without completely freeing the values of the unknowns, but the task of freeing the unknowns is far more difficult in some situations (in particular geometric ones) than in others.

Returning to some of the other items of McLeay shown above, item 6 (Test A) is similar to the Algebra item 18ii (L+M+N=L+P+N) which was described as requiring the notion of generalised number, and it is interesting to note that when changed to a "definite" form, and one in which m and t are not equal, it becomes very much easier (item 8, Test A). On the other hand it is puzzling that the other versions of item 6 (item 6, Test S and item 8, Test S) are substantially more difficult, particularly as item 6, Test S is of exactly the same form as item 15, test A and this is again easier. Possibly this is due to some undetected difference in the samples taking the two tests.

In summary, this chapter has discussed the value of extending the Algebra items and how this might be done (illustrated by items from Firth and from Ekenstam and Nilsson respectively), and has considered studies that throw further light on the use of brackets (Kieran) and, in particular, on the categories letter as object (Galvin and Bell and Clement), letter not used (Shiu), and letter as generalised number and variable (Harper and McLeay).

Chapter 15 IMPLICATIONS FOR TEACHING

It is clear from the Algebra research, and from the CSMS mathematics research generally, that

1. children's understanding of mathematical concepts and the methods they use are often quite different from what they have been taught at school, and that
2. children's understanding is often at a much lower level than is required to cope with what they are being taught.

There is a tendency to see mathematics teaching as a task of initiating children into a body of rules and procedures (which are often very powerful and for this very reason attractive to teachers). However, even if these are potentially accessible to children, it is likely that the initiation will not be entirely successful unless note is taken of finding 1., ie unless teachers take as their starting point the concepts and methods of the children themselves. This is well expressed by Case, who states that

> "until one understands what students do spontaneously, one will not be able to demonstrate the limits of this approach to them. Furthermore, until one does demonstrate the limits of whatever approach students use spontaneously, they will not thoroughly understand the necessity for using the approach to be taught"(Case,1978a,p433).

Kieran makes a similar point (see Chapter 14) when she suggests that children's adherence to the left-right rule for evaluating strings of arithmetical operations stems from their not having met situations which demonstrate that the rule is inadequate. Similarly Bell, in reviewing research into "meaningful teaching", argues that "the most successful methods, and the most plausible theory, use 'cognitive conflict'", whereby the teacher "presents a problem, allows the learner to use his natural approach, but arranges for him to reach a contradiction"(Bell, Costello and Küchemann,1980,p vii). The strategy of inducing cognitive conflict was also written into the proposal (Küchemann,1978b) of the current "Strategies and Errors" programme at Chelsea. An example of how this might be applied in the teaching of algebra would be to use letters as objects (which is how the letters are commonly interpreted, particularly by children at the Algebra levels 1 and 2, and also how the letters are often presented* when children are first introduced to algebraic manipulation) followed by a deliberate attempt to provoke conflict by presenting a context that involves objects but where the letters are to be used to represent numbers (rather than making this shift in usage surreptitiously, as commonly occurs).

With respect to finding 2., evidence from a survey undertaken by the science wing of CSMS (Shayer et al,1976; Shayer and Wylam, 1978) indicates that up to at least the age of 16 years the majority of secondary school children are at the stage of concrete operations. A similar picture emerges from the Algebra research, in that the majority of 2nd, 3rd and 4th year children were found to be at levels 1 and 2 on the Algebra test (see Chapter 6). This, and the small improvement in performance between the 3rd and 4th year children suggests that for many children there is a gross mis-match between their level of understanding of algebra and the cognitive demand of what they are taught.

Children at levels 1 and 2 may be able to solve simple equations (letter evaluated) and cope with items where the letters can be regarded as objects but they have difficulty in operating on letters as specific unknowns (and in handling the

* See Appendix 15.

ideas of generalised number and variable). Thus, for example, while they might be able to solve an equation like x+5=8, and perhaps even 24-5x=9 (see below), they are likely to have difficulty with 2x+5=x+8, where it is desirable first to manipulate the unknowns (subtract x from both sides). The mis-match referred to above raises the question of whether ways can be found of making algebra (and in particular the use of specific unknowns) more meaningful for these children, both in the sense of reducing the cognitive demand of the subject to match the level of understanding of the children (for example by trying to devise some kind of concrete referent or embodiment for the equation 2x+5=x+8) and in the sense of advancing the children's level of understanding (for example by using letters as objects and then inducing conflict, as described earlier). Research programmes currently under way at Chelsea and Nottingham should provide useful information on how, in both senses, this can be done. At the same time, the possibility has to be faced that the extent to which algebra can be made more meaningful, especially to children at levels 1 and 2, may be quite limited, and indeed a closer examination of some of the issues suggests this is the case.

Before pursuing this further, it is proposed to look at another way in which the term "meaningful" can be applied to school algebra, namely in the sense of "purposeful" or "useful". It may be the case that the difficulties of children at levels 1 and 2 stem in part from their never having been convinced of the need for algebra in solving mathematical tasks. Unfortunately, it is not easy to devise school algebra problems that can not be solved equally well by other means. Consider for example the "linear programming" question below (taken from SMP Book 3).

Fig 15.1 Linear programming question from SMP Book 3.

1. A factory makes cricket bats and tennis rackets. A cricket bat takes 1 hr. of machine time and 3 hr. of craftsman's time, while a tennis racket takes 2 hr. of machine time and 1 hr. of craftsman's time. In a day the factory has available no more than 28 hr. of machine time and 24 hr. of craftsman's time.

(*a*) If the factory makes x bats and y rackets on a particular day, write down two orderings satisfied by x and y based on: (i) machine time, (ii) craftsman's time.

(*b*) Represent these orderings graphically taking values of x from 0 to 28 and of y from 0 to 24.

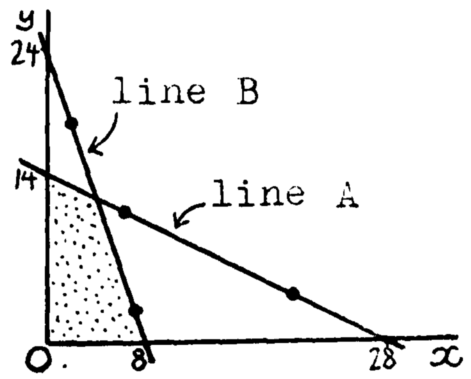
(*c*) What is the largest number of: (i) bats, (ii) rackets which could be made in a day?

(*d*) What numbers of bats and rackets must be made if the factory is to work at full capacity?

(*e*) The profits on a bat and on a racket are £1 and 10*s*. respectively. Find the maximum profit to the factory on a day when it produces: (i) only bats, (ii) only rackets, and (iii) works at full capacity.

A crucial aspect of this question (part b) is to draw the lines marked A and B in the diagram below, by finding ordered pairs such as (6,11), (20,4) and (2,18), (7,3) respectively. These can be found directly from the initial conditions given in the question, but instead children are asked to obtain them by first deriving the relationships $x+2y \leqslant 28$ and $3x+y \leqslant 24$; as well as being unnecessary this is an extremely difficult step, as the facilities of the Algebra items 17i and 22 testify.

Fig 15.2 Graphical solution of SMP linear programming question.

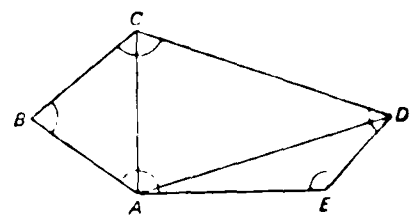


The question below is typical of questions in a chapter entitled "Letters for Numbers" in SMP Book B, whereby at the end of a series of numerical items children are asked to construct a simple algebraic expression.

Fig 15.3 "Letters for Numbers" question from SMP Book B.

How many diagonals are there from any one vertex of

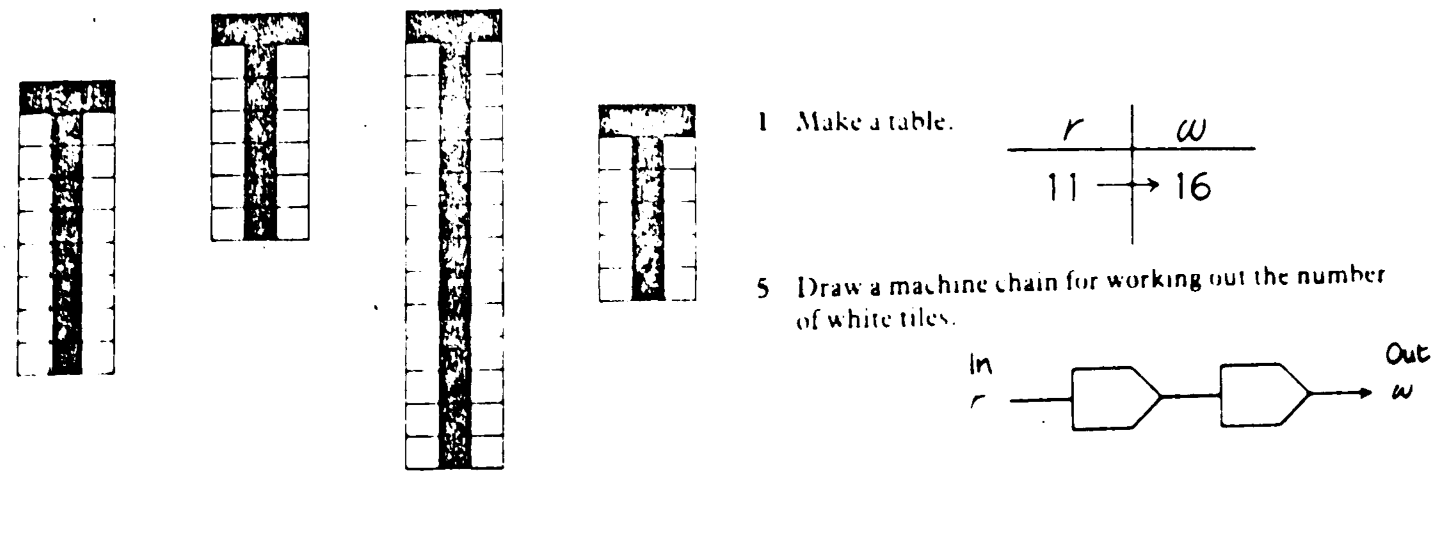
(*a*) a 5-sided polygon;
(*b*) a 6-sided polygon;
(*c*) a 7-sided polygon;
(*d*) an n-sided polygon?



Part d is almost identical to the Algebra item 15ii, which was answered correctly by only about half the 3rd year sample. On interview many children seemed bemused by the item, in part perhaps because it does not lead anywhere: in itself, what is the point of finding an expression for the number of diagonals of an n-sided or k-sided figure? Even if children were subsequently asked about the number of diagonals of a 100-sided figure, say, there is little reason why they should use the expression n-3 to find it.

It is possible to devise more purposeful examples than the one above, but generally these are also likely to be more difficult, so that it is by no means certain that children would be any more successful. For example, there would be more point to deriving an algebraic expression if the rule it describes were more complex (but therefore more difficult), since it would then be advantageous to write the rule down and to write it in a compact form. It is also likely that such rules could be formulated in several ways, in which case expressing these formulations algebraically would be a useful (though still difficult) way of demonstrating their equivalence. Such a rule is embodied in the question below (from the draft SMP 11-16 materials).

Fig 15.4 "Tiles" question from draft SMP 11-16 materials.



Here the rule for determining the number (w) of white tiles from the number (r) of red tiles could be written as $w=2(r-3)$ or $w=2r-6$, but interestingly children are instead asked to express it as a "machine chain" which in itself is likely to be easier but which makes the comparison of alternative formulations even more difficult. This illustrates the dilemma of having to choose between apparently easy but pointless or difficult but purposeful algebraic tasks.

An algebraic formulation of a rule might also be useful when the rule can be simplified. But this too is likely increase task difficulty, not just because of the required algebraic manipulation but in terms of being able to accept that the simplified expression is still equivalent to the original rule. Thus for example, when children were interviewed on the question below (which appeared on Draft 3 of the Algebra test) there were several who correctly simplified $r+s+r-s-r$ to r but then refused to use this simplification to evaluate the expression, even for $r=4001$ and $s=1903$.

Fig 15.5 Question from Draft 3.

Can you write $r + s + r - s - r$
in a simpler way?

What does $r + s + r - s - r$ stand for if $r=6$, $s=4$?
if $r=4001$, $s=1903$?

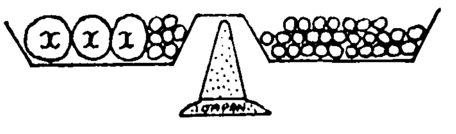
At the present time it is far from clear how important it is to children's understanding that mathematical tasks, and algebraic ones in particular, should be meaningful in the sense of "purposeful" -it may be sufficient that a task is sanctioned by the teacher ("I want you to write it as n-3; that's what we do.."). However, it is conceivable that to some children at least "pointless" tasks are just that; if in addition it is generally true that making a task more purposeful also makes it more difficult*, the possibility exists that some kinds of algebraic activity are, to all intents and purposes, inaccessible to some children.

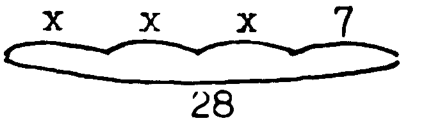
To assess how algebra might be made meaningful in the senses in which the term was initially used in this chapter, it is proposed to consider the "balance" and "line diagrams", both of which can be used to embody equations (in the manner shown below).

Fig 15.6 Use of balance and line diagram for embodying an equation.

Equation: $3x + 7 = 28$

Embodiment: Balance Line Diagram





Bell, O'Brien and Shiu (1980) found from interviews with secondary school children that their initial approach to equation solving was "usually quantitative rather than symbolic, that is by meanings rather than manipulative rules" (ibid, p124). Also, when children did try to use manipulative rules they were often unsuccessful. The table below shows three items from a written test that was given to two classes. The frequencies suggest that children could successfully apply the quantitative method even to an equation like $24-5x=9$ ("What has to be taken from 24 to leave 9", etc), but not generally to equations where the unknown is negative (items C and D). It is also likely that the children would have had difficulty applying their method to an equation like $2x+5=x+8$, where the

*A possible exception to this is the act of communicating with a computer, where even the simplest rule has to be expressed algebraically. Hart(1980), for example, found that children's performance on the Algebra test improved after being taught elementary programming.

unknown appears in several parts of the equation, even though its required value is positive.

Fig 15.7 Equation items and correct-response frequencies (O'Brien).

Item	Number of correct responses (N=53)
A. $3x + 7 = 28$	53
B. $24 - 5x = 9$	35
C. $29 = 14 - 5x$	5
D. $8x = 16 + 16x$	4

In the light of the above, a number of questions arise in considering whether introducing the notion of the balance or line diagrams would enhance children's equation solving ability:

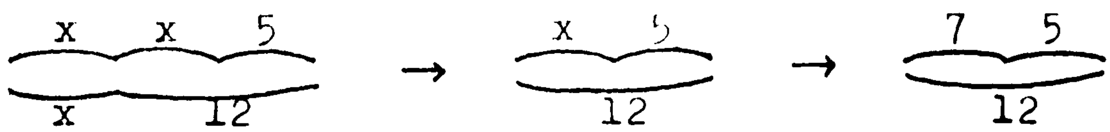
For what kind of equation are these embodiments appropriate?

For these equations, do the embodiments provide an easier method of solution than the one children naturally use?

Do the embodiments help children learn a more advanced method?

The likely answers suggest that the value of these embodiments is limited. For relatively simple equations, like A above, the embodiments <u>may</u> provide an easier method of solution, though the children's own method would seem to be adequate. The embodiments may also simplify the task of solving equations like $2x+5=x+8$, either by using the notion of taking the same amount (x) from both sides of a balance, or, in the case of a line diagram, by "matching" x:

Fig 15.8 Solution of an equation by matching, using a line diagram.



On the other hand, it is difficult to use the balance to embody equations like B, which involves subtraction, and neither embodiment is really appropriate when x is negative (C and D). Thus neither embodiment is likely to make equations of this sort substantially easier. However, this does not necessarily mean that the embodiments should be abandoned entirely (as for example Skemp (1971) suggests for the balance): for children who have reached a certain stage of cognitive development (see eg Inhelder et al,1974), it is quite possible that the use of the embodiments in this context would jolt them into a higher level of thinking -for example, by forcing them to come to terms with Harper's geometric distractor (see Chapter 14) in the case of the line diagrams. (In more general terms, Bruner, for example, would argue that a crucial aim of teaching is to help children overcome their reliance on perceptual cues, and that this is best achieved not by avoiding such cues but by demonstrating that they are misleading; a number of studies, eg Bryant 1974, lend support to this view.)

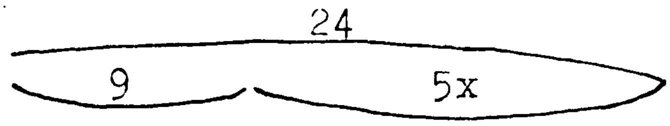
With regard to helping children understand formal methods of equation solving, the balance would seem to be more appropriate than line diagrams, since it expressly demonstrates the notion that equilibrium is maintained when the same operation is applied to both sides. However, the question needs to be studied whether children who have this notion can transfer it without difficulty from the balance to equations themselves. Bell, for example, doubts this, believing instead that

> "the attempt to characterise an equation as a balance implies a detached view of an equation which pupils find hard to coordinate with their reading of an equation as making a particular numerical statement"(Bell,1980,p8).

Put another way, manipulating certain parts of an equation while leaving the rest untouched may require a substantially greater acceptance of lack of closure than adding or subtracting matching elements from the two sides of a balance.

Line diagrams provide a very direct method of solving equations. For example, from the line diagram for $24-5x=9$

Fig 15.9 Line diagram for 24-5x=9.



it can be seen immediately that $5x=24-9$. However, it can be argued that this very directness obscures the fact that this solution (or partial solution) can be derived by applying a series of transformations to the original equation (in this case, adding 5x to both sides of the equation and then subtracting 9). In other words, because the use of line diagrams does not require children to reflect on the ways in which equations can be transformed, it is unlikely that line diagrams will help children to find or understand formal methods of solution; at best the diagrams might provide a referent against which the results of the formal methods can be checked.

Rather than using line diagrams, Petitto (1979) investigated whether the task of solving "unfamiliar" and numerically "difficult" equations like C, below, would be facilitated by presenting children with similar but much easier equations like A and B, where the solution can be seen more or less immediately.

Fig 15.10 Set of progressively more numerically-difficult equations (Petitto).

A. $\frac{1}{2} = \frac{2}{x+3}$ B. $\frac{2}{5} = \frac{4}{x+5}$ C. $\frac{14}{23} = \frac{56}{x+2}$

In general, children had considerable difficulty transferring the notions used successfully on the easier tasks to the more difficult ones, which lends support to the view expressed about line diagrams that it is difficult to identify the underlying structure of a task that can be solved in a very direct manner. (Similarly in the realm of logic, Johnson-Laird et al,1972, and

Lunzer et al,1972, found that there was very little transfer between "familiar" and "arbitrary" versions of Wason's selection task.)

In summary, it has been argued in this chapter that ways need to be found to reduce the considerable mis-match between children's level of understanding of algebra and the cognitive demand of what they are taught. For the children at levels 1 and 2 in particular, this might be done by making algebra more plausible, by using concrete embodiments and by trying to induce cognitive conflict. However, while evidence can be found to support each of these general strategies (eg Bruner,1966), it has been argued, from an examination of specific examples and in the absence of more specific data, that however carefully these strategies are put into effect they are unlikely to eliminate the mis-match entirely.

Appendix 2.1 Draft 1 of the Algebra test

CSMS

Generalisation (Draft 1)

NAME ....

CLASS .... BOY or GIRL ....

SCHOOL .... DATE OF BIRTH ....

1.

hat ⟶ head

collar ⟶ neck

sock ⟶

⟶ hand

What is the rule?

2.

1 ⟶ 3

2 ⟶ 6

3 ⟶ 9

4 ⟶ 12

:

8 ⟶ 24

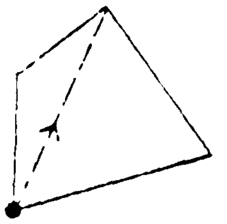
⟶ 30

300 ⟶ [illegible]

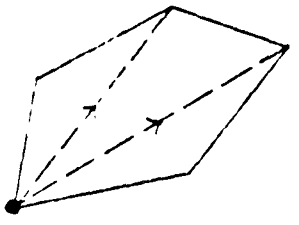
What is the rule? The first number is $\frac{1}{3}$ of the second number

Can you write it as n ⟶ 3n

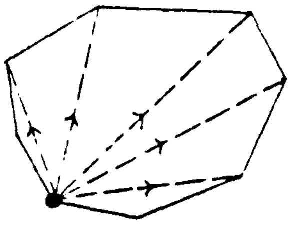
3.



For a shape with 4 sides,
starting from any corner you can draw 1 diagonal.



For a shape with 5 sides,
starting from any corner you can draw 2 diagonals.



For a shape with 8 sides,
starting from any corner you can draw 5 diagonals.

For a shape with 10 sides,
starting from any corner you can draw __ diagonals.

For a shape with 57 sides,
starting from any corner you can draw __ diagonals.

For a shape with ___ sides,
starting from any corner you can draw 125 diagonals.

You get the number of diagonals by ....

....

For a shape with k sides,
starting from any corner you can draw __ diagonals.

4.(a) If p stands for 3 and q stands for 5,

What number does $p + q$ stand for? ....

What number does $q - p$ stand for? ....

(b) If you know b and c you can

calculate a using $a = b \times c$.

If $b = 4$ and $c = 2$, what is a? ....

(c) The formula for finding the V.A.T.

on a washing machine whose basic

price is P is $V = P \div 4$.

What is the V.A.T. on a machine

whose basic price is £80? ....

5. A rough method of converting

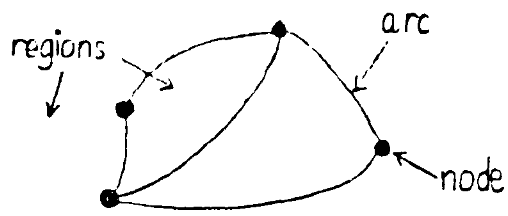
from degrees Centigrade to degrees Fahrenheit

is to multiply by 2 and add 30.

Can you write this as a formula?

....

6.



A N
7 5

A = R + N - 2

The formula relating the number of regions (r), the number of arcs (a), and the number of nodes (n) is

$$r = a - n + 2$$

R = A + 2 - N + 2
7 - 5 + 2

(i) If a = 8 and n = 5, what is r? ....

(ii) How many regions are there if there are 4 nodes and 4 arcs? ....2....

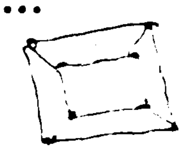
a - 4 + 2

(iii) If there are 6 regions and 4 nodes, How many arcs must there be? ....

(iv) Re-arrange the formula to give you the number of arcs, given the number of regions and the number of nodes. (Write it as a = ...)

R N A
A = R + N - 2

....



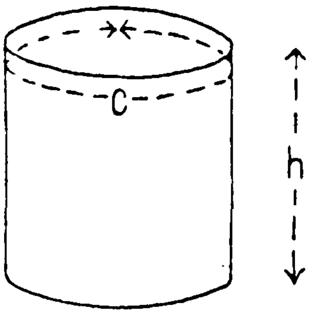
6 R
8 N
4 A

12 = 8 + 6 - 2

(v) If you have a network and you add another node and two more arcs, how many extra regions do you get? ....1....

A = R + N - 2

7.



The area of metal A used to make the curved surface of a tin can with height h and circumference c is given by the formula

$$A = h \times c$$

(i) Find A if $h = 5$ and $c = 20$ ....

(ii) If for a can of height 10 cm it requires an area of metal of 80 cm^2 to make the curved surface, what must the circumference be? ....

(iii) Rearrange the formula so that you can find the circumference given the height and the area of the curved surface. ....

(iv) If you have a can whose curved surface is made from a piece of metal of area 200 cm^2 and you want one which is twice as tall with a circumference three times as long, what area of metal would you need? ....

8. What is the relationship between x and y if

(i) $x + 5 = b$

$y + 5 = b$

X = Y

(ii) $x = a$

$y = b$

$x + a + y + b = 12$

All the letters = 3

9. When are the following true?

6 5 4 10 3 2

(i) $a + b + c = c + a + b$

15 If A + B + C have a total the same as C + A + B

(ii) $m + n + q = m + p + q$

If M + N + Q shave the same total as M + P + Q, [illegible]

10. What can you say about a if

(i) $a + b = 10$
and $a < b$

A < 5

(ii) $a + b = 10$
and $a - b = 4$

7 - 3

A > 6 <

A = 7

(iii) $a + b = 10$
and $a \times b < 21$

(iv) $a < 25$

and $a + a > 30$

Appendix 2.2 Draft 4.3 of the Algebra test

Algebra 1 (draft 4.3) **PART 2.**

Name ....

Class ....

Trial Item 1.

What number is represented by $a + 3$ if $a = 6$ ....9....

if $a = 10$ ....13....

if $a = 2$ ....5....

What number is represented by $3a$ if $a = 4$ ....12....

if $a = 7$ ....21....

Trial Item 2.

Fill in the gaps:

$$x \longrightarrow 2x$$

$1 \longrightarrow 2$

$3 \longrightarrow 6$

$4 \longrightarrow 8$

$6 \longrightarrow 12$

$9 \longrightarrow 18$

Trial Item 3.

Fill in the gaps:

$$x \longrightarrow x+1$$

$1 \longrightarrow 2$

$2 \longrightarrow 3$

$5 \longrightarrow 6$

$7 \longrightarrow 8$

$8 \longrightarrow 9$

$n \longrightarrow n+1$

1. Fill in the gaps for each of these:

$x \longrightarrow x+3$

$4 \longrightarrow 7$

$6 \longrightarrow .$

$10 \longrightarrow .$

$r \longrightarrow .$

$x \longrightarrow 5x$

$2 \longrightarrow .$

$6 \longrightarrow .$

$f \longrightarrow .$

$. \longrightarrow 20$

$2h \longrightarrow .$

$x \longrightarrow .$

$2 \longrightarrow 6$

$3 \longrightarrow 7$

$5 \longrightarrow 9$

$8 \longrightarrow .$

$y \longrightarrow .$

$t+1 \longrightarrow .$

2. Write down the smallest and largest of these:

		smallest	largest
(i)	3, 14, 4, 6, 12;	....	....
(ii)	n + 4, n + 1, n - 3, n, n - 7;	....	....

3. Add 5 onto each of these:

15	7	n	23	n + 4	3n	n - 6	n + n
......	......	......	......	......	......	......	......

4. Multiply each of these by 2:

15	7	n	23	n + 4	3n	n - 6	n + n
......	......	......	......	......	......	......	......

5. If n + 364 = 543
n + 365 = ...

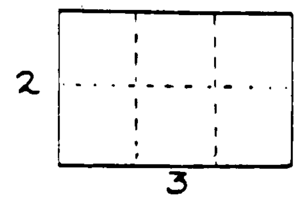
If a + b = 85
a + b + 2 = ...

If e + f = 7
e + f + g = ...

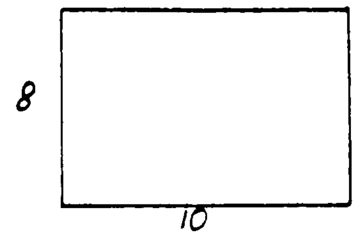
6. Complete this table:

3	2	5
4	6	10
1	8	9
2	4	.
6	e	.
d + 1	5	.
g + 4	g - 4	.
b	c	.

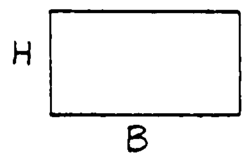
7. What are the areas of these rectangles?



Area A = ...

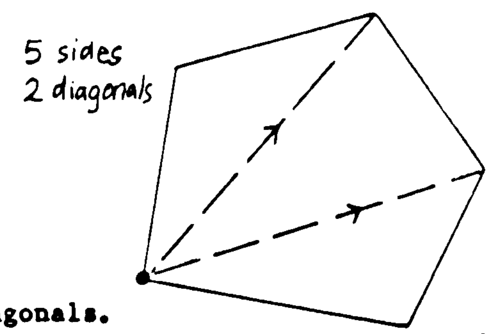


A = ...



A = ...

8. In a shape like this you can get the number of diagonals by taking 3 away from the number of sides.



If a shape had 57 sides you would get .... diagonals.

If a shape had k sides you would get .... diagonals.

9. 1 more than 4 can be written as 4 + 1.

1 more than y can be written as ......

2 less than 8 can be written as 8 - 2.

2 less than y can be written as ......

5 less than y can be written as ......

n more than y can be written as ......

10. If I have x pence and you have y pence, how much do we have altogether? ....

11. If $p + q + 4 = 83$ then $(p + 1) + (q + 1) + 4 =$ ....

12. Normally John and Mary each earn £X in a week.
How much is this altogether? ....

This week John has earned £Y more than usual and Mary has earned £Y less than usual.

How much have they earned altogether? ....

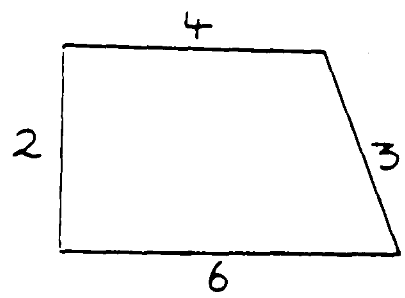
13. Can you write $(r + s) + (r - s)$
in a simpler way? ....

14. $a = b + 3.$ What is a if $b=2$? ....
What happens to a if b is increased by 2? ....

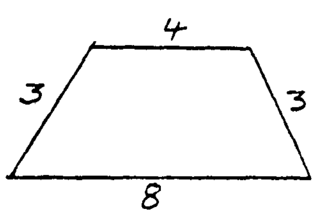
$f = 3g + 1.$ What happens to f if g is increased by 2? ....

15. What number does r stand for if $r = s + t$
and $r + s + t = 24.$ ....

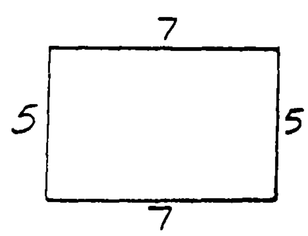
16. The perimeter of this shape
(in other words, the distance round the shape)
is equal to $6 + 3 + 4 + 2$,
which equals 15.



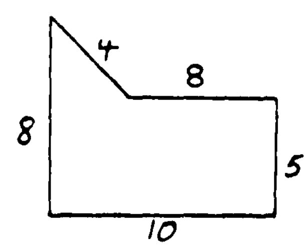
Work out the perimeter of each
of these shapes:



p = ...



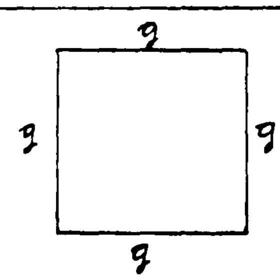
p = ...



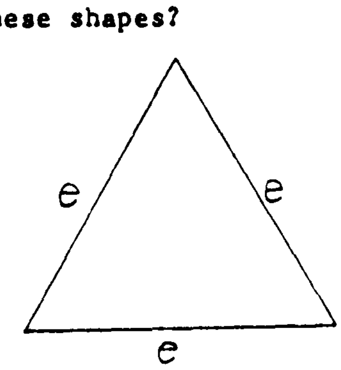
p = ...

17.

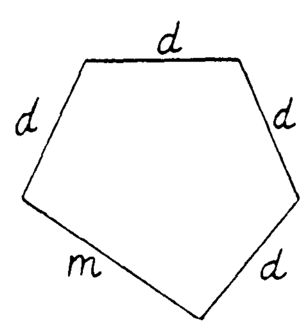
This square has sides of length g.
So, for its perimeter, we can write $p = 4g$.



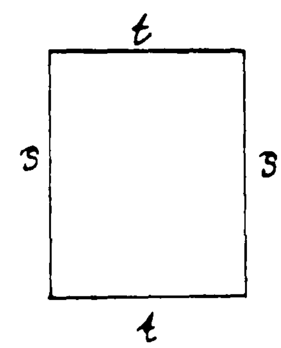
What can we write for the perimeter of each
of these shapes?



p = ...

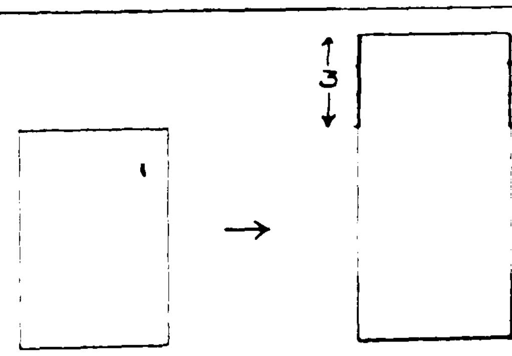


p = ...



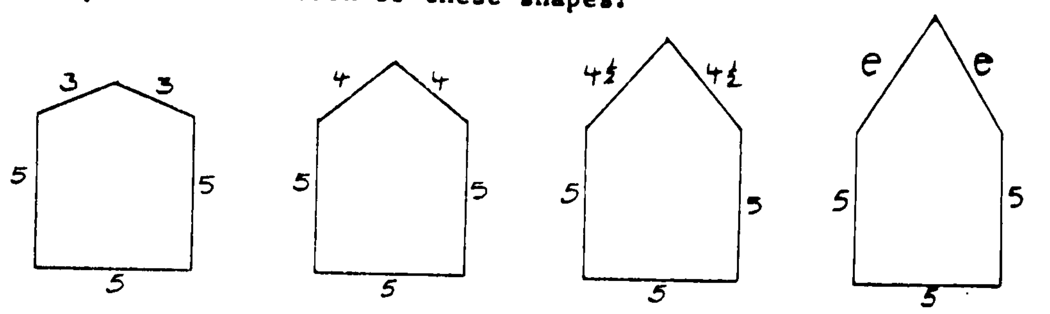
p = ...

18.



What happens to the perimeter
of this rectangle
if we make one pair of sides 3 units longer? ....

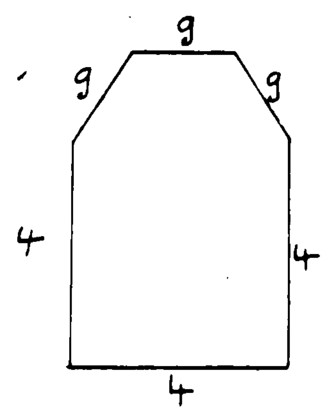
19. Find the perimeter of each of these shapes:



p = ...   p = ...   p = ...   p = ...

For the last shape,
what happens to the perimeter
if e is made 3 units longer? ....

20.

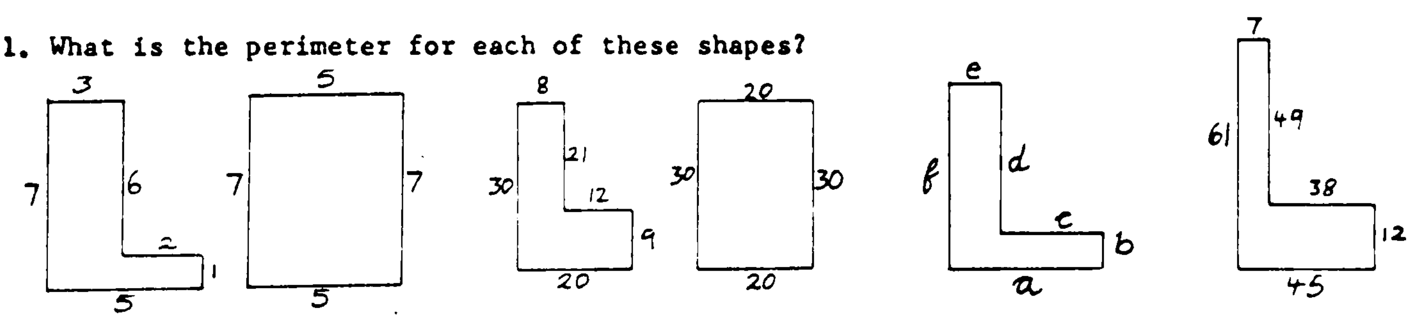


For this shape, we can write
$p = 3g + 12$.

What happens to p if g is increased by 2?

....

21. What is the perimeter for each of these shapes?



p = ...   p = ...   p = ...   p = ...   p = ...   p = ...

22. Can you write $r - s + r + s - r$ in a simpler way? ....

Now, what number does $r - s + r + s - r$ stand for

if $r = 8$, $s = 3$; ....

if $r = 3002$, $s = 1904$. ....

Appendix 2.3 Final version of the Algebra test

Algebra 1 Name ....

Trial Item 1

What number does a + 4 stand for if a = 2 ....

if a = 5 ....

What number does 4a stand for if a = 2 ....

if a = 5 ....

Trial Item 2

Fill in the gaps:
(work down the page)

$x \longrightarrow 3x$	$x \longrightarrow x+3$	$x \longrightarrow 7x$	$x \longrightarrow x+8$
$2 \longrightarrow 6$	$5 \longrightarrow 8$	$2 \longrightarrow .$	$3 \longrightarrow .$
$5 \longrightarrow .$	$4 \longrightarrow .$		
	$n \longrightarrow .$		

Algebra 1

Name .... *School* .... *Class* ......

Date .... *Date of Birth* ....
day *month* *year*

Boy or Girl ....

1. Fill in the gaps:

$x \longrightarrow x + 2$ $\qquad$ $x \longrightarrow 4x$

$6 \longrightarrow .$ $\qquad$ $3 \longrightarrow .$

$r \longrightarrow .$

2. Write down the smallest and the largest of these: **smallest** **largest**

n + 1, n + 4, n − 3, n, n − 7. ....

3. Which is the larger, 2n or n + 2 ? ....

Explain: ....

4. 4 added to n can be written as n + 4. Add 4 onto each of these:

8 $\qquad$ n + 5 $\qquad$ 3n

NOT MARKED

....

n multiplied by 4 can be written as **4n**. Multiply each of these by 4:

8 $\qquad$ n + 5 $\qquad$ 3n

NOT MARKED

....

5. If a + b = 43 $\qquad$ a + b + 2 = ...

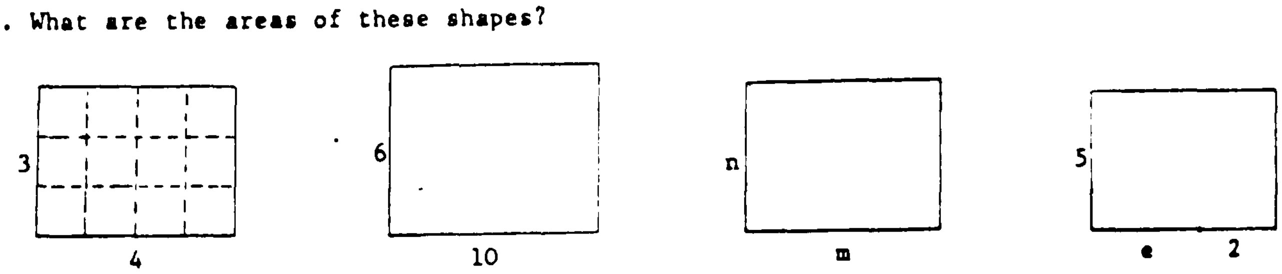
If n − 246 = 762 $\qquad$ n − 247 = ...

If e + f = 8 $\qquad$ e + f + g = ...

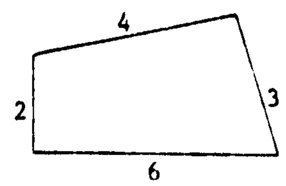
6. What can you say about a if a + 5 = 8 ....

What can you say about b if b + 2 is equal to 2b ....

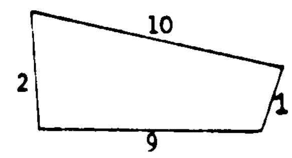
7. What are the areas of these shapes?



A = ... $\qquad$ A = ... $\qquad$ A = ... $\qquad$ A = ...

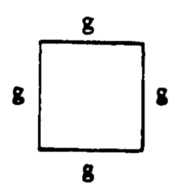


8. The perimeter of this shape is equal to $6 + 3 + 4 + 2$, which equals 15.



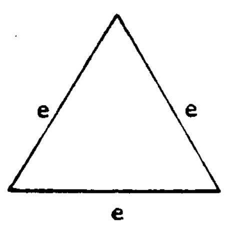
Work out the **perimeter** of this shape. $p =$ ....

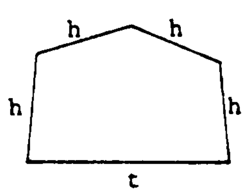
9.

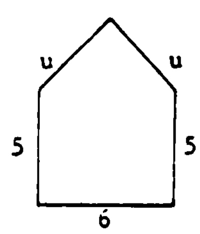


This square has sides of length g.
So, for its perimeter, we can write $p = 4g$.

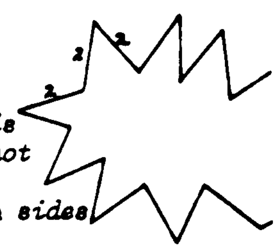
What can we write for the perimeter of each of these shapes?







Part of this figure is not drawn.
There are n *sides altogether, all of length 2.*



$p =$ .... $p =$ .... $p =$ .... $p =$ ....

---

10. Cabbages cost 8 pence each and turnips cost 6 pence each.

If c stands for the *number* of cabbages bought and t stands for the *number* of turnips bought, what does

$8c + 6t$ stand for? ....

What is the total number of vegetables bought? ....

---

11. What can you say about u if $u = v + 3$ and $v = 1$ ....

What can you say about m if $m = 3n + 1$ and $n = 4$ ....

---

12. If John has J marbles and Peter has P marbles, what could you write for the number of marbles they have altogether? ....

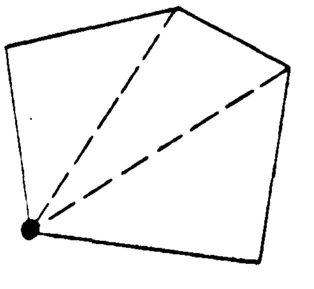
13. <u>a + 3a</u> can be written more simply as <u>4a</u>.

Write these more simply, where possible:

$2a + 5a =$	....		
$2a + 5b =$	....	$3a - (b + a) =$	....
$(a + b) + a =$	....	$a + 4 + a - 4 =$	....
$2a + 5b + a =$	....	$3a - b + a =$	....
$(a - b) + b =$	....	$(a + b) + (a - b) =$	....

14. What can you say about r if $r = s + t$
and $r + s + t = 30$ ....

15.



In a shape like this
you can work out the number of diagonals by
<u>taking away 3</u> from the number of sides.

So, a shape with 5 sides has 2 diagonals;

a shape with 57 sides has .... diagonals;

a shape with k sides has .... diagonals.

16. What can you say about c if $c + d = 10$
and c is less than d ....

17. Mary's basic wage is £20 per week.
She is also paid another £2 for each hour of overtime that she works.

If h stands for the number of hours of overtime that she works, and
if W stands for her <u>total</u> wage (in £'s)
write down an equation connecting W and h: ....

What would Mary's total wage be if she
worked 4 hours of overtime? ....

18. When are the following true -always, never, or sometimes?
Underline the correct answer:

$A + B + C = C + A + B$ Always. Never. Sometimes, when ....

$L + M + N = L + P + N$ Always. Never. Sometimes, when ....

---

19. $a = b + 3$. What happens to a if b is increased by 2? ....

$f = 3g + 1$. What happens to f if g is increased by 2? ....

---

20. Cakes cost c pence each and buns cost b pence each.
If I buy 4 cakes and 3 buns,
what does $4c + 3b$ stand for? ....

---

21. If this equation ⟶ $(x + 1)^3 + x = 349$
is true when $x = 6$,

then
what value of x
will make this equation ⟶ $(5x + 1)^3 + 5x = 349$
true?

$x =$ ....

---

22. Blue pencils cost 5 pence each and red pencils cost 6 pence each.
I buy some blue and some red pencils and altogether it costs me 90 pence.

If b is the number of blue pencils bought, and
if r is the number of red pencils bought,
what can you write down about b and r? ....

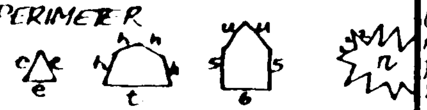
---

23. You can feed any number into this machine: $\rightarrow \boxed{\div\ 10} \rightarrow \boxed{\times\ 5} \rightarrow$

Can you find another machine that has the same overall effect? $\rightarrow \boxed{\times \ \ldots} \rightarrow \boxed{\div \ \ldots} \rightarrow$

Appendix 4.1 The Algebra test Marking Scheme

	1			2	3	4				5			6		7 AREA			
	X→X+2 6→	r→	X+4X 3→	SMALLEST/LARGEST	WHICH IS LARGER? 2n OR n+2	ADD 4 ONTO n+5	3n	MULTIPLY BY 4 n+5	3n	a+b=43 a+b+2=	e+m...	e+f=8 e+f+g=	a+5=8	b+2 ... 2b	5[ ]4	6[ ]10	n[ ]m	5[ ]e+2
1 CORRECT	8	r+2	12	n-7, n+4	"DEPENDS" + AWARE OF AT LEAST TWO OF: n=2 n=1 n=0 n NEGATIVE (n>2 ONLY)	n+9	3n+4	4(n+5) 4n+20	12n	45	761	8+g	3	2	12	60	nm n×m	5(e+2) 5e+10
2 CORRECT WEAK					"DEPENDS" + ONLY ONE OF: n=2 n=1 n=0 n NEG												MISUNDERSTOOD BUT CONSISTENT WITH	
3 AMBIGUOUS [INADEQUATE FUNCTION +, x; NOT NECESSARILY WRONG]		r2 2r x+2		OTHER WAY ROUND n+4, n-7 MISSING:	"DEPENDS" OR "USUALLY 2n" BUT NONE OF: n=2, 1, 0, NEG	n+5+4 n+54 n9	3n+4 IFF "n+5+4" 3n4 n×3+4	4n+5 n+5×4 4n5 n54 n+54	3n×4 3n4 43n			8g		"b+2=2b"			n,m	5e+2 e+2×5 5e2 e25 MISUNDERSTOOD
4 AMBIGUOUS ASSOCIATION [WRONG FUNCTION +, x BUT NOT WRONG OPERATION IN ...]		r×2			2n	n×9 n+5×4	3n×4		12+n			8×g					n+m	5e×2
5 ASSOCIATION WITH LETTER [EXPLICIT WRONG OPERATION ON ...]						n+20	7n 12n	n+20 n+9	7n									e+10 10e 7e
6 ASSOCIATION LETTER IGNORED		+2 2				9	7 12	20 9	7 12			9						10 7
7 SPECIFIC/CONSISTENT [ALPHABET CODE]		t 20 19 21 2t				23 22 24 19 LV W+9 3+9	3r 3w	4s	3r			15 ONLY						35 50
8 SPECIFIC/OBVIOUS (* ANY OTHER NUMERICAL ANS)	12	*	7	*	n+2 "SAME"	*	*	*	*	*	763	12	*	b=1 b=0	*	*	12	25 20 30
9 OTHERS		INCLUDE S		ONE OR MORE WRONG														
0 BLANK																		

2/8	9a PERIMETER				10 CABBAGES AND TURNIPS 8c + 6t	ACC VEG	11 n+3 n=...		12 MARKS
 10, 9, p=	 e	t	6	n					
22 23 21 20	3e e3	4h+t 4h+t	2u+16 u^2+16 2^u+16	2n 2^n $\frac{n}{2}$	EXPLICIT "COST OF c CABBAGES AND t TURNIPS"	c+t	4	13	J+P
	LONG e+e+e	h+h+h+h+t	u+u+u+16 2u+7.5+6	n2 n^2	LITERAL "8p×(c) CABBAGES PLUS 6p×(t) TURNIPS"				
	e.e.e e,e,e (indexes)	hhhht 4ht	uuu556 2u16	n,2 n^2 2^n 2	IMPLICIT "(TOTAL) COST" "COST OF CABBAGES AND TURNIPS"	c,t ct	V= n−3	n= $\frac{n-1}{3}$	J and P, J, P, J_1+P_0, n−n, J∧P
	e×e×e	4h×t	2u×16	n+2	"THE NUMBER of cabbages or turnips" (GENERAL)	c×t			J×P
		5h	32u 8u 2u+150 2u+60	32n 32+n 17n 6n					
					1 CAB, 1 TURNIP 8p+6p 14p	2			
	15 125			28 p	100 8 CABS 6 TURNIPS AND 100	100			26 27 25 Z
*	*	*	*	32+42 34 6 16 17	8 CAB, 6 TURNIPS (14 CABBAGES AND TURNIPS)	14	2	*	*
			ARITH ERROR	include 32+2n n×n nn	INCLUDE 14ct 14c+t 14c,t etc. 8×c + 6×c				J=P

3/13	2a + 5a = 7a	2a + 5b = 2a + 5b	(a + b) + a = 2a + b	2a + 5b + a = 3a + 5b	(a − b) + b = a	3a − (b + a) = 2a − b	a + 4 + a − 4 = 2a	3a − b + a = 4a − b	(a + b) + (a − b) = 2a	14 r = ... r + ...	15 GIVE ...
	2a + 5a	2a + 5b	(a + b) + a	2a + 5b + a	(a − b) + b	3a − (b + a)	a + 4 + a − 4	3a − b + a	(a + b) + (a − b)		
	7a	2a + 5b OR "IMPOSSIBLE" OR 'BLANK' IF AT LEAST ONE OTHER BLANK	2a + b	3a + 5b	a	2a − b	2a	4a − b $4^a - b$	2a 2^a	r = 15	54
									2a + 1·b OR a + a		
	BRACKET EXPANDED: (EVEN IF WRONG) SIMPLIFIED BUT INADEQUATE PUNCTUATION 'BLANK' IF >1 OTHER BLANK		$a^2 + ab$ $a^2 + b$ 2a + ab 2a, b 2a b a^2b	3a5b $2a^2 + 5b$	$ab - b^2$ ab − 2b b^2a	$3ab - 3a^2$ $3a^2 - b$ (2ab) 3ab − 4a	a^2	4ab $3a^2 - b$	$a^2 - b^2$ a^2	r write r = 30 − r r = 30 − r "r < 30"	
	STRAIGHT FORWARD		WRONGS 2a × b	(OR ×	a − 2b a + b	INSTEAD 4a − b	OF +) 2a + 8 0 a	2a − b 3a − b	a 2a + 2b 2a − 2b 2a + b 0		
	ASSOCIATIONS (ab) ABSORPTIONS (a) REPLACEMENTS (c) 8a	10b 12a	3ab 4a 2c	8ab	3ab ab b	5ab 2b		5ab 2ab 2a			
	*	*	*	*	*	*	*	*	*	10	*
			INCLUDE ITEM WRITTEN DOWN UNCHANGED (with or without brackets) OR NOT COMPLETELY SIMPLIFIED ab+a abn		ab + b a − 4b		2a + 4 − 4				
		'BLANK' IF ALL OTHERS BLANK (other than 2a+5a = 7a)									

[illegible] k →	16 → c + d = 10 c bluest band	17 → WAGES	4 HOURS	18 A+B+C = C+A+B	L+M+N = L+P+N	19 a = b+3 Δb = 2 Δa = ?	f = 3g+1 Δg = 2 Δf = ?	20 CAKES AND BUNS 4c + 3b	21 x = [illegible]	22 BLUE AND RED PENCILS	[illegible]	Code
k − 3	c < 5 (c > 5)	W = 20 + 2h w = 20 + 2 × h	28	ALWAYS OR ALWAYS WHEN "GENERAL" "POSITIVE"	SOMETIMES (ALWAYS) WHEN M = P	Δa = 2	Δf = 6	EXPLICIT "TOTAL COST" "COST OF 4 CAKES AND 3 BUNS"	6/5 1·2	5b : 0r = 90	×5 +50	1 CORRECT
WRONG BUT CONSISTENT WITH (k − 2 ONLY)	c ≤ 5 > c ≤ 4 c ≤ 4 c ≤ 6 SYSTEMATIC LIST (4 VALUES) 1 4 1234 (0 1 2 3 4 5)	TWO NUMERICAL PAIRS OF VALUES (CORRECT) w = 28 h = 4 w = 30, h = 5 etc OR 28, 4 30, 5 etc				STATIC — a + 2 = b + 5 a + 2 = b + 3 a = b + 1 a + 2.	"CORRECT" f + 6 = ANSWER f = 3(g+2) + 1 (f = 3g + 7) f = 3g 5 f + 6.	LITERAL "4 CAKES AT 4p EACH AND 3 (BUNS) AT 4p EACH"		TWO NUMERICAL PAIRS OF VALUES (CORRECT) FROM 6 10 12, 5 15 p 0, 15 OR b = 3, r < 8		2 CORRECT WEAK [SEVERAL CORRECT NUMERICAL ANSWERS]
3k (−2) 3 − k	C = 10 − d.	"w + 2h = wages" w + 2h 20 + 2h EXPRESSIONS INVOLVING "nw, nh" (CORRECT) AMBIGUOUS PAIR eg W = 20 + h/2		ALWAYS / SOMETIMES WHEN "SPECIFIC" eg H = 1 B = 2 C = 3	SOMETIMES. ALWAYS (= NUMBERED) SOMETIMES WHEN M = NUMBER ? eg	"NO CHANGE" (NOT "NOTHING") AMBIGUOUS	PUNCTUATION f = 3g + 2 + 1 f = g + 2 + 3	IMPLICIT "COST" "AMOUNT"	1 < x < 2 x = 5/6 (NOT x = 1 x = 2)	THE LETTERS STAND FOR CARDINALITY OF PENCILS b + r = NUMBER OF PENCILS EXPRESSIONS INVOLVING "b, r" (CORRECT)	"NO"	3 AMBIGUOUS
k3 (k2) 3k (2k) k + 3 (k+2) k × 3 (k×2)	UNSYSTEMATIC 2 1 3 4, etc OR LESS THAN 4 VALUES (MORE THAN ONE)	w = 20 + h h → w 20 + h w = w + h w = wages + h TOTAL WAGE = W + h				"INCREASES" OR "INCREASES BY ..." eg "f increases by 2" f + 2		"NUMBER OF CAKES + BUNS" (GENERAL)		b + r = amount of money b + r = 90 b = 90 b + r ≠ 90		4 LETTER AS OBJECT GENERAL.
		W + h. wh w∩h				STATIC — INCORRECT a = b + 5 a = 2b + 3 a = 3b + 3 ALL "OTHER" EXPRESSIONS in b	f = 3g + 3 f = 3g − 1 f = 6g			b + r		5 LETTER AS / ASSOCIATION OBJECT / GENERAL
−3 (−2) 3 (2)	C = 4 5 3 2 1 0	ONE PAIR OF VALUES (CORRECT) w = 28 h = 4 etc OR 28, 4 etc				a = 2	f = 6	(1 bun, 1 cake) 4p + 3p 7p		ONE PAIR OF VALUES (CORRECT) b = 6 r = 10 etc [ACCEPT "r = 10"]		6 ONE CORRECT NUMERICAL ANSWER
k → h (g) i (h) 8 (i) 9 (k)		2h = 24w 4h = 23w, etc OR 24 = w + 2h 25 = w + 4h etc (NUMBERS CORRECT)				(a → c) etc	(f → l) etc	25p (NOT ≥ 18p)		6b + 3r etc (NUMBERS CORRECT) b = 30 r = 60 etc (6 BLUE + 10 RED)	CONSISTENT 10 NOT 5 NOT ANY 5 NUMBER	7 LETTER AS OBJECT SPECIFIC / CONSISTENT
*	* (ALL ARITH ERRORS OTHER THAN C = 5)	w = 20, h = 2 (IMPLIES w = 20 h = 2) OR 20w + 2h £22 OR w = 10h, etc 22 = w + h	*			a = 5	f = 7	4 CAKES, 3 BUNS 7 cakes	30	FIFTEEN b = [illegible] r = 6 b < r b + r = [illegible] etc	5 10 OR 10 5	8 LETTER AS OBJECT SPECIFIC / OBVIOUS
K	INCLUDE c < d	INCLUDE ALL ARITH ERRORS eg w = 28 h = 8 w + h > 20 25 = w + 8 2h = 4w 25w + 8h h/2 w/20		NEVER	NEVER	"CHANGES" a ≠ b + 3 a < EXPRESSION	f ≠ 3g + 1 f < EXPRESSION	INCLUDE 7cb 7c, b 7c + b etc 4xc + 3xb		(b = PENCILS cost) INCLUDE ALL ARITHMETICAL ERRORS eg b = 10 r = 6	INCLUDE 5 NOT 50 5 NOT 10	9 OTHERS
					—							0 BLANK

Appendix 4.2 Frequency of each Response-Type and Facility of Algebra items (%), for 1976 sample (including 103 above average 5th year children)

year / item	code 1				code 2				code 3				code 4				code 5			
	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
1	83	88	88	99		0														
	65	75	68	99					4	1	4	1	0		0			0		
	74	86	90	98					0									0		
2	56	72	72	86						2	2	1	2							
3	2	3	1	22	2	3	4	10	2	2	2	4	27	16	6	[illegible]				
4	61	68	69	90					6	4	6	3	0				1	1	2	
	21	36	41	78					4	3	4	3	0	0		1	41	31	30	12
	8	17	25	51		0			12	19	18	33	0	0		1	39	31	29	10
	39	45	50	64					6	5	8	12	1	0	0	1	11	14	13	17
5	92	97	95	100																
	60	74	74	91					0											
	25	41	50	82					6	3	7				0				0	
6	86	92	93	100					0		0									
	29	41	40	79		0			3	2	3	1								
7	81	91	90	100					0											
	79	89	90	99		0			0											
	51	67	75	99	3	1	1		1	0	0		5	5	3		0			
	7	12	16	49	0	0			15	18	18	21	4	3	2		27	28	31	1
8	95	97	96	99	0		0													
9	90	94	93	98	0	0			0		0									
	57	68	72	96	0	0	1		22	20	13		1	1	3	2	4	1	2	
	28	37	45	69	26	27	22	21	20	16	12	1	1	2	3	3	2	3	2	3
	9	23	23	46	14	15	18	34	1	1	1	1	0	1	1	2	3	5	4	1
10		0	2	3	1	1	1	1	3	4	5	21	2	3	2	2				
	2	4	5	20					0	0	0	1			0					
11	49	61	70	93	0				0	1	0			0						
	44	62	[illegible]	91						0	0									
12	[illegible]	[illegible]	[illegible]	[illegible]					18	14	16	13	2	1	3	1				
13	77	86	87	97					0	0	0	2	1	0	0		10	8	6	1
	29	45	51	87					7	10	6	1	1	0	1		45	34	34	11
	38	53	55	75					9	10	18	15	1	1	1	1	19	15	8	2
	40	60	66	94					3	2	4	2	1	1	1		26	20	17	3
	15	23	32	62		0	0		3	4	7	10	25	28	22	18	20	15	15	4
	11	15	20	52					3	4	6	11	20	24	28	18	22	19	16	4
	30	44	52	82	0	1	1		3	2	4	3	16	13	12	5	11	11	9	3
	27	47	56	86					2	1	2	1	6	5	6	4	24	20	14	3
	11	12	25	54	1	2	1		3	5	9	[illegible]	20	29	24	12	10	[illegible]	[illegible]	[illegible]
14	27	35	31	52					2	6	8	2								
15	63	75	72	90		0														
	33	51	53	88	1	1	1		2	4	3	1	1	1	0	1	0			
16	7	11	16	32	14	19	19	33	1	4	4	6	1	1	1					
17	2	5	8	31		0			1	3	3	4	6	13	19	22	13	14	12	4
	65	77	76	97			0													
18	58	72	73	96					14	11	12	3			0					
	11	25	27	63					16	14	12	9								
19	9	21	26	51	1	10	7	18	1	2	1		5	5	6	4	22	20	18	15
	3	7	13	26	1	2	3	11	1	2	2	1	10	15	17	31	29	22	23	15
20	2	11	11	7	4	3	4	9	6	8	9	[illegible]	2	4	2	10				
21	4	13	16	[illegible]	0				1	2	4	2								
22	2	16	12	52	0	1	1		1	1	1	2	31	17	17	16	2	2	3	
23	2	3	3	5					1	2	2	2								

year / item	code 6				code 7				code 8				code 9				code 0				Facility % code 1+2			
	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
1			0						5	4	4		6	5	6		6	3	3	1	83	88	88	99
	7	9	6		3	1	0		5	4	3		6	4	6		10	6	13		65	75	68	99
		0	0						8	4	3		10	8	10	1	9	4	9	1	74	86	90	98
2									4	5	3		26	20	20	1	2	2	3	2	[illegible]	[illegible]	72	86
3					0	0			24	16	15	9	1	1	2		7	4	5		[illegible]	6	10	22
4	20	20	14	4	0		0		3	2	2		4	3	3	3	4	3	4		61	68	69	90
	17	16	12	4					5	5	3		5	5	5	2	6	3	5		22	36	41	78
	16	15	11	3		0			7	5	3		9	6	5	2	10	6	9	1	8	17	25	51
	16	16	11						7	7	3		9	7	7	2	11	6	9	2	39	45	50	64
5									3	2	2		1				4	1	3		92	97	95	100
									22	13	11	7	10	8	7	2	9	5	8		59	74	74	91
	10	6	6	3	4	2	1	3	29	26	16	5	14	11	8	4	13	10	13	4	25	41	50	82
6									2	1	1		2	2	1		11	5	5		86	92	93	100
									11	12	13	5	27	23	21	6	31	22	24	10	29	41	40	79
7									12	5	5		0	1	0		6	3	4		81	91	90	100
									14	8	6		0	1	0		6	3	4	1	79	89	90	99
					0				6	5	3		14	10	6	1	19	11	12		54	68	76	99
	10	13	9	6	1	0	0	1	6	5	5	1	12	9	7	2	19	13	12	5	7	12	16	49
8									2	1	1	1		0			3	1	3		95	[illegible]	96	99
9					1	0	0		2	2	1		2	2	1	1	4	2	5	1	91	94	93	98
									3	2	1		7	5	3		5	3	6	1	58	68	73	96
		0							4	3	1	1	11	7	7	1	8	5	8	1	54	64	67	90
	0		0		3	0	0		25	18	17	1	20	16	15	5	23	22	22	11	24	38	41	80
10	5	5	2	6	20	23	37	27	49	52	38	16	2	7	8	[illegible]	9	5	6	1	1	1	3	6
	6	5	3	5	9	7	8		[illegible]	[illegible]	[illegible]	[illegible]	8	6	4	3	10	6	8	1	2	4	5	20
11		0							15	14	7	1	14	10	5		22	14	17		[illegible]	[illegible]	[illegible]	[illegible]
									22	14	8		7	5	3		27	19	21		[illegible]	[illegible]	[illegible]	[illegible]
12					6	2	3		7	3	3		6	5	3		15	11	11		[illegible]	[illegible]	[illegible]	[illegible]
13									5	2	2		0	0	0		7	3	4		77	86	87	97
									6	4	2		1	1	0		11	7	6	2	29	45	51	86
									7	5	3		5	4	3	3	22	13	12	5	38	53	55	75
									7	5	3		2	1	0		20	11	10	1	40	60	66	94
									7	5	3		4	5	5	1	25	20	16	5	15	23	32	62
									8	4	2		5	7	6	4	32	27	23	12	11	15	20	52
									9	6	3		5	7	4	2	25	16	16	3	30	44	52	82
									8	5	1		3	2	1		31	20	20	4	27	47	56	86
									7	3	1		5	5	7	5	4	17	22	[illegible]	11	19	26	54
14									24	21	21	8	18	16	16		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	19	77
15			0						23	16	16	6		0	0		15	9	11	4	[illegible]	[illegible]	[illegible]	90
	3	1	1		14		8	8	11	9	5		14	10	10	[illegible]	22	15	19	4	[illegible]	[illegible]	[illegible]	88
16	43	39	35	10					2	1	1		9	[illegible]	9	[illegible]	23	16	16	7	21	30	35	70
17	0	0			1	2	1		14	11	11	10	21	19	16	14	33	23	11	6	2	5	[illegible]	31
									16	11	11	2	4	3	3		16	9	11	1	[illegible]	[illegible]	76	97
18													10	[illegible]	4		17	10	11	1	[illegible]	[illegible]	[illegible]	96
										0			[illegible]	[illegible]	50	27	18	11	11	1	[illegible]	[illegible]	[illegible]	63
19	1	0	1		2	1	1		8	8	9	2	15	15	11	4	26	18	19	2	9	21	26	51
	2	3	2		1	0	0		2	2	4	1	4	15	13	6	[illegible]	[illegible]	[illegible]	[illegible]	3	7	13	26
20	2	2	2		1	1	1	1	36	39	26	14	19	10	21	4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	23	[illegible]
21									2	4	2	3	43	35	30	13	[illegible]	[illegible]	[illegible]	[illegible]	4	12	16	39
22			2	3	5	5	4	6	4	4	4	2	27	23	23	9	45	[illegible]	32	21	2	10	13	52
23					25	21	24	29	30	35	29	35	3	1	2	1	40	37	40	23	[illegible]	[illegible]	[illegible]	[illegible]

Appendix 5 Cross-Tabulation of Children's Algebra Levels by Response-Codes, for Items 9i, 6i, 11i, 9ii, 13iv, 5iii, 4ii, 4iii, 22

Item 9i (level 1 item)

Level of Children

	Level 0	Level 1	Level 2	Level 3	Level 4	Number of Children
Code 1+2	32	94	100	99	100	
Other Codes	68	6	0	1	0	
	100%	100%	100%	100%	100%	
Number of Children	(202)	(1118)	(662)	(752)	(189)	(2923)

Item 6i (level 1 item)

Level of Children

	Level 0	Level 1	Level 2	Level 3	Level 4	
Code 1	30	88	99	99	100	
Other Codes	70	12	2	1	0	
	100%	100%	100%	100%	100%	
Number of Children	(202)	(1118)	(662)	(752)	(189)	(2923)

Item 11i (level 2 item)

Level of Children

	Level 0		Level 1		Level 2		Level 3		Level 4			
Code 1	1	9	18	28	32	86	38	89	11	97	100%	(1749)
Code 3	0	0	42	0	25	1	33	1	0	0	100%	(12)
Code 8	5	9	67	22	12	7	15	7	1	2	100%	(365)
Code 9	11	15	71	18	11	5	6	2	1	1	100%	(285)
Code 0	27	67	70	32	2	2	1	1	0	0	100%	(509)
		100%		100%		100%		100%		100%		
Number of Children		(202)		(1116)		(661)		(752)		(189)		(2920)

Item 9ii (level 2 item)

Level of Children

	Level 0	Level 1	Level 2	Level 3	Level 4		Number of Children
Code 1	1 7	24 41	31 89	35 89	10 100	100%	(1922)
Code 2	0 0	40 1	47 1	7 0	7 1	100%	(15)
Code 3	5 12	78 38	9 7	9 6	0 0	100%	(538)
Code 4	0 0	44 2	17 1	39 3	0 0	100%	(52)
Code 5	7 2	83 5	5 1	5 0	0 0	100%	(60)
Code 8	53 15	47 2	0 0	0 0	0 0	100%	(58)
Code 9	21 16	67 9	6 1	7 1	0 0	100%	(152)
Code 0	75 47	23 3	0 0	2 0	0 0	100%	(126)
	100%	100%	100%	100%	100%		
Number of Children	(202)	(1118)	(662)	(752)	(189)		(2923)

Item 13iv (level 2 item)

Level of Children

	Level 0	Level 1	Level 2	Level 3	Level 4		
Code 1	0 2	14 20	31 75	43 92	12 99	100%	(1606)
Code 3	1 1	85 7	10 1	3 0	0 0	100%	(87)
Code 4	0 0	62 2	21 1	17 1	0 0	100%	(29)
Code 5	6 17	73 41	15 14	7 6	0 1	100%	(623)
Code 8	35 25	59 7	5 1	1 0	0 0	100%	(14[illegible])
Code 9	7 1	58 2	32 2	3 0	0 0	100%	(31)
Code 0	27 55	60 22	11 7	2 1	0 0	100%	(405)
	100%	100%	100%	100%	100%		
Number of Children	(202)	(1118)	(662)	(752)	(189)		(2923)

Item 5iii (level 3 item)

Level of Children

	Level 0		Level 1		Level 2		Level 3		Level 4			Number of Children
Code 1	0	2	10	10	18	30	55	82	16	95	100%	(1119)
Code 3	8	5	59	8	29	6	5	1	0	0	100%	(147)
Code 6	7	8	65	13	21	7	7	2	0	0	100%	(219)
Code 7	5	2	46	3	28	3	20	2	1	1	100%	(76)
Code 8	6	19	55	34	32	33	7	7	0	1	100%	(694)
Code 9	13	21	57	17	24	12	6	2	1	1	100%	(325)
Code 0	26	44	49	15	16	8	7	3	2	3	100%	(341)
		100%		100%		100%		100%		100%		
Number of Children		(202)		(1117)		(662)		(751)		(189)		(2921)

Item 4ii (level 3 item)

Level of Children

	Level 0		Level 1		Level 2		Level 3		Level 4			
Code 1	1	3	10	8	13	20	58	75	19	96	100%	(971)
Code 3	5	3	32	5	36	6	24	4	4	2	100%	(109)
Code 4	0	0	60	0	0	0	40	0	0	0	100%	(5)
Code 5	5	26	51	46	31	47	13	17	0	1	100%	(996)
Code 6	13	27	57	22	27	18	4	2	1	1	100%	(441)
Code 8	16	10	58	7	24	5	2	0	0	0	100%	(127)
Code 9	11	8	63	8	17	4	8	2	1	1	100%	(149)
Code 0	39	24	49	6	8	2	4	1	0	0	100%	(125)
		100%		100%		100%		100%		100%		
Number of Children		(202)		(1118)		(662)		(752)		(189)		(2923)

Item 4iii (level 4 item)

Level of Children

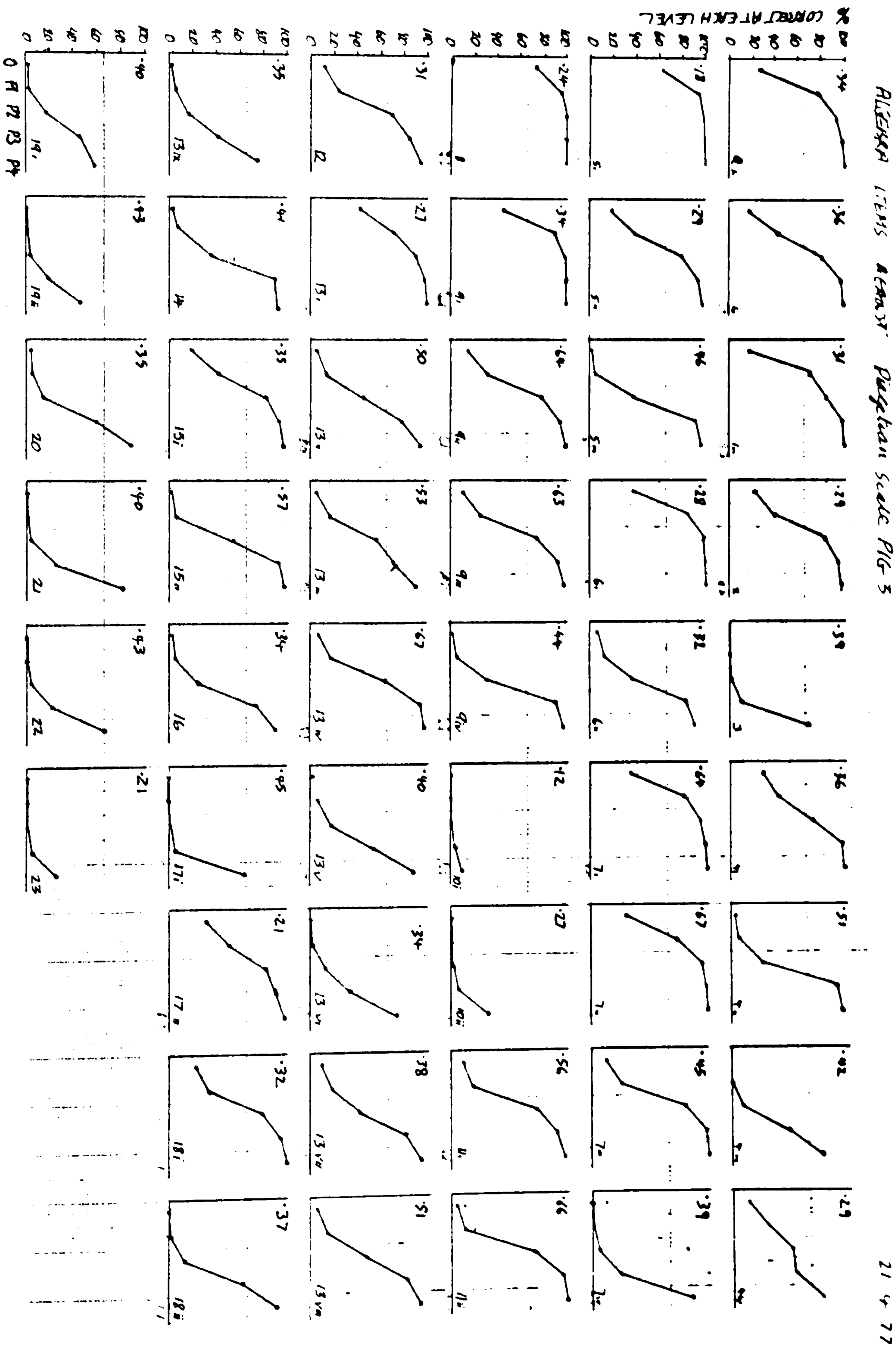
	Level 0		Level 1		Level 2		Level 3		Level 4			Number of Children
Code 1	0	1	5	2	13	9	47	30	35	88	100%	(478)
Code 3	0	1	10	4	23	16	63	40	5	12	100%	(477)
Code 4	0	0	0	0	25	0	75	0	0	0	100%	(4)
Code 5	3	16	50	43	30	44	17	22	0	0	100%	(963)
Code 6	11	22	58	21	28	18	3	2	0	1	100%	(415)
Code 8	20	14	58	8	20	4	1	0	0	0	100%	(144)
Code 9	13	12	54	10	18	5	16	4	0	0	100%	(197)
Code 0	28	34	57	12	10	4	5	2	0	0	100%	(243)
		100%		100%		100%		100%		100%		
Number of Children	(202)		(1117)		(662)		(751)		(189)			(2921)

Item 22 (level 4 item)

Level of Children

	Level 0		Level 1		Level 2		Level 3		Level 4			
Code 1	0	0	2	0	7	3	39	14	52	73	100%	(264)
Code 2	0	0	14	0	0	0	29	1	57	4	100%	(14)
Code 3	0	0	19	0	26	1	48	2	7	1	100%	(27)
Code 4	0	1	19	7	25	16	52	29	4	10	100%	(424)
Code 5	0	0	48	2	42	3	11	1	0	0	100%	(48)
Code 6	0	0	22	2	33	4	38	4	6	3	100%	(81)
Code 7	1	1	33	5	30	7	33	7	3	3	100%	(162)
Code 8	5	3	31	3	30	5	33	5	1	1	100%	(121)
Code 9	7	24	50	31	22	23	20	18	1	5	100%	(697)
Code 0	13	72	50	49	23	37	14	19	0	1	100%	(1085)
		100%		100%		100%		100%		100%		
Number of Children	(202)		(1118)		(662)		(752)		(189)			(2923)

Appendix 8 Swags for Algebra PIG 3 Scale, and Mean Phi (RMS) for each item with every other item



Appendix 9 Phi values between Algebra items (derived from total 1976 sample, N=2923)

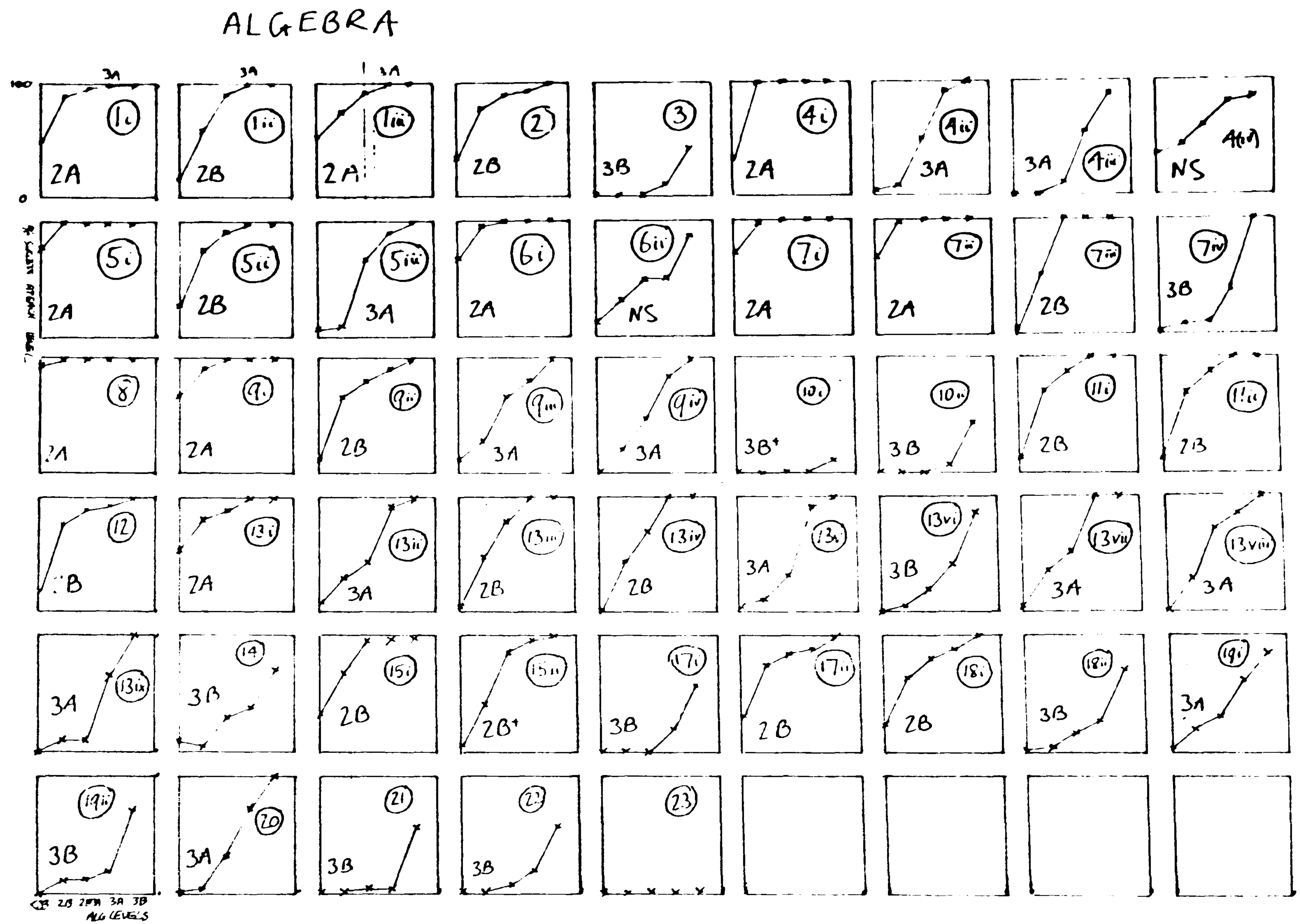
EASY HARD

•	22	43	27	21	24	17	26	15	19	14	14	24	16	21	10	20	19	10	31	15	20	18	16	16	10	14	08	14	12	14	11	12	09	11	09	01	01	08	09	08	08	07	06	06	06	05	04	03	04	05
	•	18	28	16	19	20	21	13	12	13	21	21	25	19	17	16	20	16	22	16	23	20	16	20	11	15	15	14	17	15	15	13	15	14	10	11	12	10	09	10	10	08	08	08	07	07	06	05	09	10
		•	32	29	33	22	31	19	24	25	24	31	22	21	21	36	29	34	22	22	27	27	24	23	17	20	18	20	18	18	17	18	16	15	13	14	13	12	13	12	11	10	09	09	08	07	06	05	09	07
			•	30	34	23	31	20	26	29	24	28	32	28	19	28	30	27	30	18	33	30	26	27	15	24	20	23	17	11	13	21	19	19	16	14	17	14	13	13	14	11	12	11	09	08	08	05	05	03
				•	80	20	25	24	20	28	21	27	25	24	17	25	34	24	27	26	31	18	23	26	15	25	20	22	23	22	19	23	19	19	18	15	15	16	15	15	14	13	13	12	11	11	09	07	06	04
					•	26	30	24	23	32	25	31	28	28	20	28	40	27	30	27	35	30	24	31	17	28	23	24	26	24	21	25	25	11	20	16	18	16	17	17	14	15	14	13	12	11	10	07	07	05
						•	22	48	18	24	44	26	22	22	20	21	25	22	24	21	27	25	22	31	12	21	19	20	23	23	21	22	18	18	17	15	16	13	13	14	14	14	13	13	11	10	10	06	07	05
							•	20	25	25	22	28	23	28	21	28	32	28	28	24	30	38	32	27	17	29	27	31	24	23	20	21	23	20	17	17	19	18	15	16	17	13	14	13	11	11	11	05	08	04
								•	21	21	38	26	19	20	17	23	27	25	26	19	29	25	22	30	14	24	20	21	26	25	23	23	23	18	18	18	18	16	17	17	16	16	16	14	13	12	11	08	08	05
									•	22	21	30	26	31	17	24	26	24	28	21	32	28	24	31	15	26	22	24	23	26	23	23	21	23	19	21	22	15	16	16	16	17	18	17	13	13	13	08	09	06
										•	28	30	31	31	28	29	36	30	32	24	38	35	33	38	18	33	27	29	31	31	26	29	30	28	26	20	22	22	20	22	21	19	18	16	14	16	14	09	08	05
											•	29	28	27	29	30	39	31	31	30	36	35	28	43	10	30	25	29	34	28	27	32	30	27	24	25	26	20	21	24	20	20	21	19	16	16	15	11	10	05
												•	30	35	23	28	36	28	35	31	40	35	28	50	18	33	25	28	30	32	28	30	27	29	25	21	26	19	21	22	18	17	18	18	16	16	15	12	09	06
													•	30	25	30	34	29	39	31	39	35	29	37	18	32	27	29	33	32	31	31	34	26	25	22	25	21	20	23	23	19	20	18	16	17	15	11	11	04
														•	22	28	37	29	34	31	41	40	34	41	14	36	27	33	33	34	27	35	30	31	27	25	31	24	23	23	21	19	21	20	11	16	17	12	11	07
															•	24	32	24	24	11	31	29	26	33	12	21	25	25	30	30	23	30	42	24	26	19	23	10	17	25	11	21	19	17	16	16	16	10	10	05
																•	34	18	31	38	35	43	37	38	15	34	19	35	36	30	29	31	33	28	23	24	24	20	21	25	22	23	20	18	17	19	17	10	12	08
																	•	34	42	38	49	48	38	52	22	45	37	37	46	36	32	42	38	33	30	28	30	24	26	30	26	26	24	25	19	20	18	13	11	07
																		•	35	36	38	45	40	40	17	35	30	34	36	33	29	31	34	31	26	26	26	21	22	26	22	24	21	20	18	20	18	13	12	09
																			•	34	24	43	35	46	15	42	33	36	39	41	36	38	35	33	28	27	32	24	27	27	24	24	24	22	19	20	19	13	12	08
																				•	39	35	32	42	15	34	25	30	38	30	31	34	32	28	24	26	28	22	21	24	22	23	21	19	18	20	18	15	11	06
																					•	50	42	54	21	48	38	44	46	48	41	44	42	40	33	33	37	28	29	32	29	28	28	26	22	22	20	15	14	07
																						•	69	48	18	63	45	65	45	39	33	40	43	37	35	31	36	34	29	32	33	28	27	27	20	22	20	15	14	09
																							•	41	14	51	39	54	37	35	28	34	35	32	36	27	30	37	23	26	34	26	23	22	18	24	21	16	15	09
																								•	20	49	38	43	55	47	40	52	49	42	38	36	42	31	32	36	30	31	34	32	25	28	25	19	18	11
																									•	19	19	15	18	16	11	18	10	13	17	14	12	20	13	31	17	17	14	17	12	16	15	07	14	04
																										•	44	52	46	41	31	43	45	37	38	35	38	38	27	25	33	32	31	31	24	25	25	17	16	11
																											•	43	39	37	31	35	41	32	43	31	37	43	25	35	35	33	30	29	24	28	24	16	17	10
																												•	40	38	31	37	41	35	37	29	36	36	27	34	35	31	30	28	23	26	24	17	16	04
																													•	43	36	47	49	39	39	37	41	31	29	40	34	36	36	33	25	30	28	19	20	01
																														•	39	42	45	43	37	34	40	31	30	35	33	32	32	31	26	29	27	19	16	12
																															•	38	37	32	31	30	34	23	25	27	27	30	30	25	24	28	26	18	16	07
																																•	46	38	40	39	40	33	33	38	32	37	38	36	29	32	32	20	21	12
																																	•	40	42	40	42	35	31	45	37	36	37	39	28	33	29	21	20	10
																																		•	39	33	40	30	27	37	32	32	35	29	16	29	27	16	19	10
																																			•	35	38	41	28	39	41	40	39	37	29	18	31	19	15	13
																																				•	37	31	30	37	31	36	41	40	32	36	35	18	22	14
																																					•	31	29	34	33	37	40	33	27	33	35	22	24	11
																																						•	21	37	37	32	33	34	26	32	36	24	22	12
																																							•	34	20	28	30	31	30	27	24	16	20	06
																																								•	35	44	42	43	32	40	38	25	29	11
																																									•	37	34	32	26	38	36	24	25	06
																																										•	44	43	35	44	43	24	30	17
																																											•	43	37	42	44	28	29	16
																																												•	37	40	53	37	33	17
																																													•	38	36	25	27	08
																																														•	44	30	37	13
																																															•	41	34	18
																																																•	25	30
																																																	•	11
																																																		•

Appendix 10 Correspondence between Variable Numbers and Item Numbers

Item Number	Variable Number
1i	1
1ii	2
1iii	3
2	4
3	5
4i	6
4ii	7
4iii	8
4iv	9
5i	10
5ii	11
5iii	12
6i	13
6ii	14
7i	15
7ii	16
7iii	17
7iv	18
8	19
9i	20
9ii	21
9iii	22
9iv	23
10i	24
10ii	25
11i	26
11ii	27
12	28
13i	29
13ii	30
13iii	31
13iv	32
13v	33
13vi	34
13vii	35
13viii	36
13ix	37
14	38
15i	39
15ii	40
16	41
17i	42
17ii	43
18i	44
18ii	45
19i	46
19ii	47
20	48
21	49
22	50
23	51

Appendix 12.1 Shayer's Swags for Algebra items



Appendix 12.2 The Pendulum task

TASK III

SCIENCE REASONING TASKS

NAME ....

BOY OR GIRL ....

SCHOOL ....

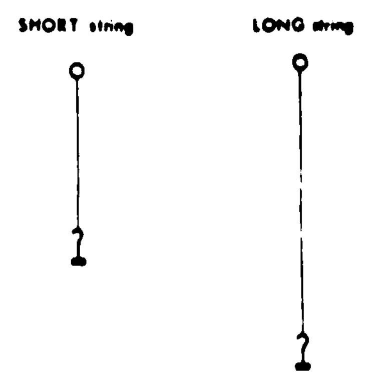
TODAY'S DATE ....

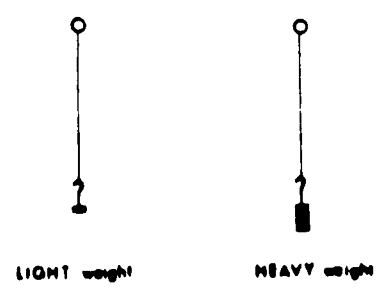
CLASS ....

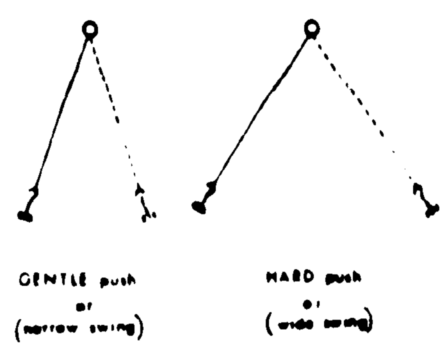
DATE OF BIRTH ....
day month year

THE PENDULUM

We are going to make a pendulum, using a SHORT or LONG string,
and a LIGHT or HEAVY weight,
and we will exert a GENTLE or HARD push,







This task will be about the number of swings the pendulum makes in a given time (½ – minute).

1

length	weight	push	number of swings in ½-minute

A.1 SHORT string, HEAVY weight, GENTLE push.

Your guess:__________swings. Experiment 1 →

A.2 LONG string, LIGHT weight, GENTLE push.

Your guess:__________swings. Experiment 2 →

A.3 What effect do you think LENGTH, WEIGHT, and PUSH have on the number of swings in half a minute?

LENGTH:

WEIGHT:

PUSH.

A.4a Now what can we tell, if anything, just from these experiments, about the effect of LENGTH, WEIGHT and PUSH on the number of swings?

LENGTH:

A.4b WEIGHT:

PUSH:

A.4c Write down one more experiment that you think would be worth trying next, and explain why you have chosen it. Also explain how this new experient ties in with experiment 1 or 2:

	length	weight	push	
A.5 Imagine that we start again with experiment 1 →	SHORT	HEAVY	GENTLE	
Which other arrangements would you use to test the effect that LENGTH has on the number of swings? → (But please use as few arrangements as possible; put a star (*) next to any arrangements that you don't really need.)				
A.6 Again starting with experiment 1 → how would you test for the effect that WEIGHT has? → (But, again, use as few arrangements as possible; put a star (*) next to any arrangements that you don't really need.)	SHORT	HEAVY	GENTLE	
A.7 Imagine someone tried these two arrangements (with another pendulum) →	long	heavy	hard	15
	short	heavy	gentle	20
a. What do they tell us about the effect of the PUSH?				
b. If there are any other arrangements that you think you would really need to be sure of the effect of the push, write them down → (and cross-out any of the original two arrangements that you don't need).				

		length	weight	push	number of swings in ½-minute
B.1	Experiment 1 →	SHORT	HEAVY	GENTLE	
B.2	Experiment 2 →	LONG	LIGHT	GENTLE	
B.3	LONG string, HEAVY weight, HARD push. your guess: ________ swings. Experiment 3 →				
B.4	SHORT string, LIGHT weight, GENTLE push. your guess: ________ swings. Experiment 4 →				

B.5 Now write down what these four experiments alone tell us about the effect of **LENGTH, WEIGHT and PUSH** on the number of swings,

and for each factor, note down only those experiments that you need to use:

a. LENGTH: b. experiments

c. WEIGHT: d. experiments

e. PUSH: f. experiments

g. Is the evidence weaker for deciding about one of the factors than it is for the others? ____________

If so, say which factor: ____________

and

EITHER show that the evidence is still sufficient,

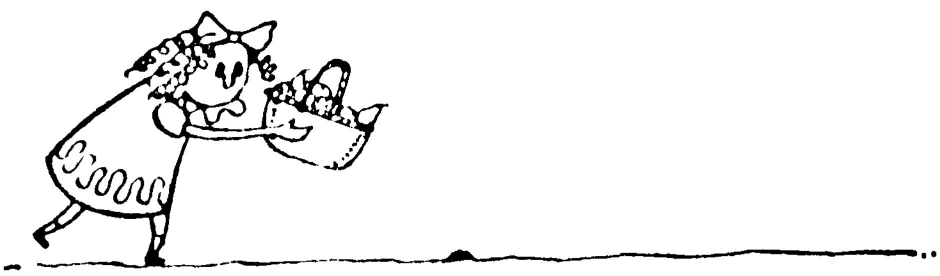
OR explain why it is insufficient.

Appendix 15 Use of Letters as Objects (in an Intoductory Chapter for First Year Secondary School Children)

ALGEBRA

CHAPTER 14

USE OF LETTERS



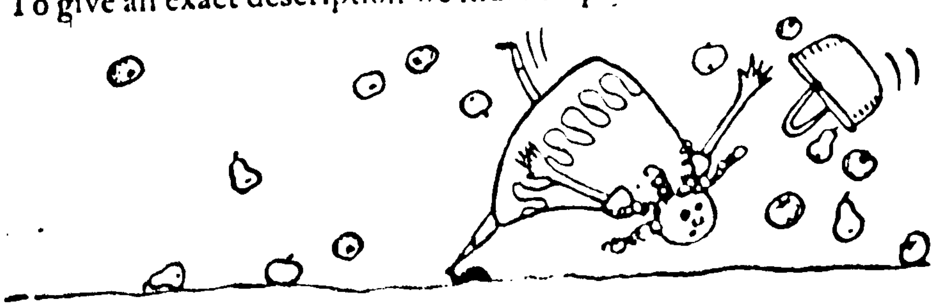
This basket of fruit can be described in more detail by saying:

(a) The basket contains 3 pounds of fruit.

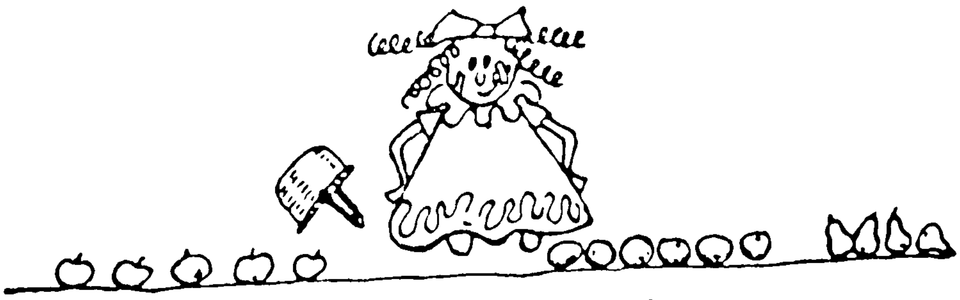
(b) The basket is full of apples, pears and oranges.

(c) The basket contains 3 pounds of apples, pears and oranges.

To give an exact description we must empty the basket,



sort out the different kinds and count them.



Five apples + six oranges + four pears.

This can be written $5a + 6o + 4p$.

151

From Shaw and Wright (1970).

Bibliography

Adey P: 1980, The development of a science reasoning task. In Archenhold et al, 1980.

Adi H: 1978, Intellectual development and reversibility of thought in equation solving. Journal for Research in Mathematics Education, 9, 204-213.

APU: 1980a, Mathematical Development: Primary Survey Report No 1. London: HMSO.

APU: 1980b, Mathematical Development: Secondary Survey Report No 1. London: HMSO.

Archenhold WF, Driver RH, Orton A and Wood-Robinson C (Ed): 1980, Cognitive Development Research in Science and Mathematics: Proceedings of an International Seminar, September 1979. The University of Leeds.

Bart W and Krus D: 1973, An ordering-theoretic method to determine hierarchies amoung items. Educational and Psychological Measurement, 33, 291-300.

Bell AW: 1976, The Learning of General Mathematical Strategies. Unpublished Ph D thesis, University of Nottingham.

Bell AW: 1980, Research and development on the teaching of algebra. Working Paper, Shell Centre for Mathematical Education, University of Nottingham.

Bell AW, Costello J and Küchemann DE: 1980, A Review of Research in Mathematical Education; Section A: Research on Learning and Teaching. Shell Centre for Mathematical Education, University of Nottingham.

Bell AW, O'Brien D and Shiu C: 1980, Designing teaching in the light of research on understanding. In Proceedings of the 4th International Conference for the Psychology of Mathematics Education, August 1980. Berkeley, California.

Bloom BS, Englehart MD, Furst EJ, Hill WH and Krathwohl DR: 1965, Taxonomy of Educational Objectives; Handbook 1: Cognitive Domain. London: Longmans.

Brainerd CJ: 1978, Learning research and Piagetian theory. In Brainerd S: Alternatives to Piaget. London: Academic Press.

Brown G and Desforges C: 1977, Piagetian psychology and education: time for revision. Br J educ Psychol, 47, 7-17.

Brown M and Küchemann DE: 1976, Is it an add, Miss? Mathematics in School, 5, n5 and 6, n1.

Bruner JS: 1966, Toward a Theory of Instruction. Cambridge, Massachusetts: The Belknap Press.

Bryant PE and Trabasso T: 1971, Transitive inferences and memory in young children. Nature, 232, 456-458.

Calvert B: 1958, Non-Verbal Test DH. NFER.

Carrol JB: 1961, The nature of the data, or how to choose a correlation coefficient. Psychometrika, 26, 347-372.

Case R: 1974, Structures and strictures: some functional limitations on the course of cognitive growth. Cognitive Psychology, 6, 544-573.

Case R: 1978a, A developmentally based theory and technology of instruction. Review of Educational Research, 48, 439-463.

Case R: 1978b, Intellectual development from birth to adulthood: a neo-Piagetian interpretation. In Siegler, 1978.

Chilton RJ: 1966, Computer generated data and the statistical significance of scalogram. Sociometry, 29, 175-181.

Clement J, Lochhead J and Monk GS: 1980a, Translation difficulties in learning mathematics. Technical Report, Cognitive Development Project, University of Massachusetts, Amherst. (To appear in American Mathematical Monthly.)

Clement J, Lochhead J and Soloway E: 1980b, Positive effects of computer programming on students' understanding of algebra. Paper presented at the national (USA) meeting of the Association for Computing Machinery, October 1980.

Collis KF: 1972, A Study of Concrete and Formal Reasoning in School Mathematics. Unpublished Ph D thesis, University of Newcastle, NSW.

Collis KF: 1974, Cognitive development and mathematics learning. London: PMEW, Chelsea College.

Collis KF: 1975a, A Study of Concrete and Formal Operations in School Mathematics: A Piagetian Viewpoint. Hawthorn, Victoria: Australian Council for Eduactional Research.

Collis KF: 1975b, The Development of Formal Reasoning: Report of a SSRC sponsored project carried out at the University of Nottingham School of Education during 1974. University of Newcastle, NSW.

Comrey AL and Levonian E: 1958, A comparison of three point coefficients in Factor analyses of MMPI items. Educational and Psychological Measurement, 18, 739-755.

Craik FIM and Lockhart RS: 1972, Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Crannell CW and Parish JM: 1957, A comparison of immediate memory span for digits, letters and words. Journal of Psychology, 44, 319-327.

CSMS: 1975, Concepts in Secondary School Mathematics and Science. Research proposal submitted to the Social Science Research Council.

CSMS: 1981, Children's Understanding of Mathematics: 11 - 16 (CSMS Mathematics Team). London: Murray.

Davidoff MD and Goheen HW: 1953, A table for the rapid determination of the tetrachoric correlation coefficient. Psychometrika, 18, 115-121.

De Boysson-Bardies B and O'Regan K: 1973, What children do in spite of adults' hypotheses. Nature, 246, 531-534.

Desforges C and Brown G: 1979, The educational utility of Piaget: a reply to Shayer. Br J educ Psychol, 49, 277-281.

Donaldson M: 1978, Children's Minds. London: Fontana.

Ekenstam A Af and Nilsson M: 1979, A new approach to the assessment of children's mathematical competence. Educational Studies in Mathematics, 10, 41-66.

Ennis RH: 1975, Children's ability to handle Piaget's propositional logic: a conceptual critique. Review of Educational Research, 45, 1-41.

Entwistle NJ: 1979, Stages, levels, styles or strategies: dilemmas in the description of thinking. Educational Review, 31, 123-132.

Ferguson GA: 1941, The factorial interpretation of test difficulty Psychometrika, 6, 323-333.

Firth DE: 1975, A Study of Rule Dependence in Algebra. Unpublished M Phil thesis, University of Nottingham.

Furneaux WD and Rees R: 1978, The structure of mathematical ability. British Journal of Psychology, 69, 507-512.

Galvin WP and Bell AW: 1977, Aspects of difficulty in the solution of problems involving the formation of equations. Shell Centre for Mathematical Education, University of Nottingham.

Gelman R: 1969, Conservation acquisition: a problem of learning to attend to relevant attributes. Journal of Experimental Child Psychology, 7, 167-187.

Gelman R: 1972, Logical capacity of very young children: number invariance rules. Child Development, 43, 75-90.

Green BF: 1956, A method of scalogram analysis using summary statistics. Psychometrika, 21, 79-88.

Guilford JP: 1965, Fundamental Statistics in Psychology and Education. New York: McGraw-Hill.

Guttman L: 1944, A basis for scaling qualitative data. American Sociological Review, 9, 139-150.

Guttman L: 1950, The basis for scalogram analysis. In Stouffer SA, et al, Measurement and Prediction. Princeton NJ: Princeton University Press, pp 60-90.

Halford GS: 1978, An approach to the definition of cognitive developmental stages in school mathematics. Br J educ Psychol, 48, 298-314.

Halford GS and Galloway W: 1977, Children who fail to make transitive inferences can remember comparisons. Australian Journal of Psychology, 29, 1-5.

Harper E: 1978, Sketch for a theory of confusion in mathematics learning. Int J Math Educ Sci Technol, 9, 231-237.

Harper E: 1979, The Child's Interpretation of a Numerical Variable. Unpublished Ph D thesis, University of Bath.

Harper E: 1980, The boundary between arithmetic and algebra: conceptual understanding in two language systems. Int J Math Educ Sci Technol, 11, 237-243.

Hart K: 1980, Secondary School Children's Understanding of Mathematics: A Report of the Mathematics Component of the CSMS Programme. London: Centre for Science Education, Chelsea College.

Hart K: 1981, The research of CSMS. In CSMS,1981.

Hart ML: 1980, The Nottingham Programming in Mathematics Project: Report No. 1. Department of Mathematical Sciences, Trent Polytechnic, Nottingham.

Huttenlocher J and Burke D: 1976, Why does memory span increase with age? Cognitive Psychology, 8, 1-31.

Inhelder B and Piaget J: 1958, The Growth of Logical Thinking from Childhood to Adolescence. London: Routledge and Kegan Paul.

Inhelder B, Sinclair H and Bovet J: 1974, Learning and the Development of Cognition. London: Routledge and Kegan Paul.

Janvier C: 1978, The Interpretation of Complex Cartesian Graphs: Studies and Teaching Experiments. Unpublished Ph D thesis, University of Nottingham.

Johnson-Laird PN, Legrenzi P and Legrenzi MS: 1972, Reasoning and a sense of reality. British Journal of Psychology, 63, 395-400.

Johnson-Laird PN and Wason PC: 1970, A theoretical analysis of insight into a reasoning task. Cognitive Psychology, 1, 134-148.

Kaltenhauser J and Lee Y: 1976, Correlation coefficients for binary data in factor analysis. Geographical Analysis, 8, 305-313.

Keats JA: 1955, Formal and Concrete Thought Processes. Princeton: Educational Testing Service.

Kerslake DM: 1977, The understanding of graphs. Mathematics in School, 6,n2.

Kieran C: 1979, Children's operational thinking within the context of bracketing and the order of operations. In Proceedings of the Third Annual Conference of the International Group for the Psychology of Mathematics Education, July 1979. University of Warwick.

Kieran C: 1980, The interpretation of the equal sign: symbol for an equivalence relation vs. an operational symbol. In Proceedings of the 4th International Conference for the Psychology of Mathematics Education, August 1980. Berkeley, California.

Kingsley PR and Hagen JW: 1969, Induced versus spontaneous rehearsal in short-term memory in nursery school children. Developmental Psychology, 1, 40-46.

Knifong JD: 1974, Logical abilities of young children: two styles of approach. Child Development, 45, 78-83.

Küchemann DE: 1978a, Secondary school children's understanding of reflection and rotation. Osnabrücker Schriften zur Mathematik, Reihe D, Band 1: Proceedings of the Second Annual Conference of the International Group for the Psychology of Mathematics Education, September 1978. University of Osnabrück.

Küchemann DE: 1978b, Strategies and Errors. Research proposal submitted by the CSMS mathematics team to the Social Science Research Council.

Küchemann DE: 1979, Task III, The Pendulum (test and teachers' guide). Windsor: NFER.

Küchemann DE: 1980, Children's difficulties with single reflections and rotations. Mathematics in School, 9, n2.

Küchemann DE: in press, Algebra (test and teachers' guide). Windsor: NFER.

Lawson AE: 1976, M-space: is it a constraint on conservation reasoning ability? Journal of Experimental Child Psychology, 22, 40-49.

Loevinger J: 1947, A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs, 61, n4.

Loevinger J: 1948, The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. Psychological Bulletin, 45, 507-530.

Lovell K: 1980, The relevance of cognitive psychology to science and mathematics education. In Archenhold et al,1980.

Lunzer EA: 1973, Formal reasoning: a re-appraisal. Keynote paper presented at syposium of Jean Piaget Society, Philadelphia, Spring 1973. (Also published by PMEW, Chelsea College, 1976, and in Floyd A (Ed),1978, Cognitive Development in the School Years. An Open University Reader. London: Croom Helm).

Lunzer EA, Bell AW and Shiu CM: 1976, Numbers and the world of things. Shell Centre for Mathematical Education, University of Nottingham.

Lunzer EA, Harrison C and Davey M: 1972, The four-card problem and the generality of formal reasoning. Quarterly Journal of Experimental Psychology, 24, 326-339.

Malpas AJ: 1974, Objectives and cognitive demands of the School Mathematics Project's main school course. Mathematics in School, 3, n5 and 3, n6.

Malpas AJ and Brown M: 1974, Cognitive demand and difficulty of GCE O-level mathematics pretest items. Br J educ Psychol, 44, 155-162.

McLaughlin GH: 1963, Psycho-logic: a possible alternative to Piaget's formulation. Br J educ Psychol, 33. 61-67.

McLeavy H: 1980, Some Aspects of Children's Interpretations of of Literal Expressions. Unpublished M Ed Dissertation, University of Bath.

McNemar Q: 1969, Psychological Statistics. New York: John Wiley.

Menzel H: 1953, A new coefficient for scalogram analysis. Public Opinion Quarterly, 17, 268-280.

Murray FB: 1972, Acquisition of conservation through social interaction. Developmental Psychology, 6, 1-6.

O'Brien TC: 1973, Logical thinking in college students. Educational Studies in Mathematics, 5, 71-80.

Parsons C: 1960, Inhelder and Piaget's "Growth of Logical Thinking": a logician's viewpoint. British Journal of Psychology, 51, 75-84.

Pascual-Leone J: 1976, On learning and development, Piagetian style: 1. a reply to Lefebvre-Pinard. Canadian Psychological Review, 17, 270-288.

Pascual-Leone J and Smith J: 1969, The encoding and decoding of symbols by children: a new experimental paradigm and a neo-Piagetian model. Journal of Experimental Child Psychology, 8, 328-355.

Petitto A: 1979, The role of formal and non-formal thinking in doing algebra. Journal of Children's Mathematical Behavior, 2, n2, 69-82.

Rosnick P and Clement J: 1980, Learning without understanding: the effect of tutoring strategies on algebra misconceptions. Technical Report, Cognitive Development Project, University of Massachusetts, Amherst.

Ruddock GJ: 1980, Children's Understanding of Vectors and Matrices. Unpublished Ph D thesis, University of London, Chelsea College.

Shaw HA and Wright FE: 1970, Discovering Mathematics: A Course For Secondary Schools; Volume 1. 2nd (Decimal) Edition. London: Edward Arnold.

Shayer M: 1978a, A Test of the Validity of Piaget's Construct of Formal Operational Thinking. Unpublished Ph D thesis, University of London, Chelsea College.

Shayer M: 1978b, Parametric coefficients and the Yule measures. CSMS working paper.

Shayer M: 1978c, The analysis of science curricula for Piagetian level of demand. Studies in Science Education, 5, 115-130.

Shayer M: 1979, Has Piaget's construct of formal operational thinking any utility? Br J educ Psychol, 49, 265-276.

Shayer M, Küchemann DE and Wylam H: 1976, The distribution of Piagetian stages of thinking in British middle and secondary school children. Br J educ Psychol, 46, 164-173.

Shayer M and Wylam H: 1978, The distribution of Piagetian stages of thinking in British middle and secondary school children II: 14-16 year-olds and sex differentials. Br J educ Psychol, 48, 62-70.

Sheppard JL: 1974, Compensation and combinatorial systems in the acquisition and generalisation of conservation. Child Development, 45, 717-730.

Shiu C: 1978, The Development of some Mathematical Concepts in School Children. Unpublished Ph D thesis, University of Nottingham.

Siegel S: 1956, Non-Parametric Statistics for the Behavioral. Sciences. New York: McGraw-Hill.

Siegler R (Ed): 1978, Children's Thinking: What Develops? Hillsdale, N Jersey: Lawrence Erlbaum.

Skemp RR: 1971, The Psychology of Learning Mathematics. Harmondsworth: Penguin Books.

SPSS: 1975, Nie NHM, Hull CJ, Jenkins JG, Steinbrenner K and Bent DH: Statistical Package for the Social Sciences (2nd Edition). New York: McGraw-Hill.

Versey J: 1978, Scalogram analysis and cognitive development: evidence from a longitudinal study. Br J educ Psychol, 48, 71-78.

Wagner S: 1979, Mathematical variables and verbal "variables": an essential difference. Paper posted at the 3rd Annual Conference of the International Group for the Psychology of Mathematics Education, July 1979. University of Warwick.

Wason PC: 1966, Reasoning. In Foss B (Ed), New Horizons in Psychology. Harmondsworth: Penguin Books.

Wason PC and Shapiro D: 1971, Natural and contrived experience in a reasoning problem. Quarterly Journal of Experimental Psychology, 23, 63-71.

Yule GU: 1912, On the methods of measuring association between two attributes. Journal of the Royal Statistical Society, 75, 576-642.

Zimmerman BJ and Rosenthal TL: 1974, Conserving and retaining equalities and inequalities through observation and correction. Developmental Psychology, 10, 260-268.

Siegler R (Ed): 1978, Children's Thinking: What Develops? Hillsdale, N Jersey: Lawrence Erlbaum.

Skemp RR: 1971, The Psychology of Learning Mathematics. Harmondsworth: Penguin Books.

SPSS: 1975, Nie NH, Hull CJ, Jenkins JG, Steinbrenner K and Bent DH: Statistical Package for the Social Sciences (2nd Edition). New York: McGraw-Hill.

Versey J: 1978, Scalogram analysis and cognitive development: evidence from a longitudinal study. Br J educ Psychol, 48, 71-78.

Wagner S: 1979, Mathematical variables and verbal "variables": an essential difference. Paper posted at the 3rd Annual Conference of the International Group for the Psychology of Mathematics Education, July 1979. University of Warwick.

Wason PC: 1966, Reasoning. In Foss B (Ed), New Horizons in Psychology. Harmondsworth: Penguin Books.

Wason PC and Shapiro D: 1971, Natural and contrived experience in a reasoning problem. Quarterly Journal of Experimental Psychology, 23, 63-71.

Yule GU: 1912, On the methods of measuring association between two attributes. Journal of the Royal Statistical Society, 75, 576-642.

Zimmerman BJ and Rosenthal TL: 1974, Conserving and retaining equalities and inequalities through observation and correction. Developmental Psychology, 10, 260-268.